Department of Signal Processing and Acoustics

Cortical processing of the periodicity of speech sounds

Santeri Yrttiaho

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 144/2011

Cortical processing of the periodicity of speech sounds

Santeri Yrttiaho

Doctoral dissertation for the degree of Doctor of Philosophy to be presented with due permission of the School of Electrical Engineering for public examination and debate in Auditorium S1 at the Aalto University School of Electrical Engineering (Espoo, Finland) on the 20th of January 2012 at 12 noon.

Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics

Supervisor
Professor Paavo Alku

Instructors
Docent Hannu Tiitinen
Docent Patrick May

Preliminary examiners
Professor Matti Hämäläinen, Harvard Medical School, United States
Professor Eric Schröger, University of Leipzig, Germany

Opponent
PD Dr. med. Alexander Gutschalk, University of Heidelberg, Germany

Aalto University publication series
DOCTORAL DISSERTATIONS 144/2011

ISBN 978-952-60-4443-9 (printed)
ISBN 978-952-60-4444-6 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Unigrafia Oy
Helsinki 2011

Finland

The dissertation can be read at http://lib.tkk.fi/Diss/

441 697
Printed matter

Author
Santeri Yrttiaho

Name of the doctoral dissertation
Cortical processing of the periodicity of speech sounds

Publisher School of Electrical Engineering

Unit Department of Signal Processing and Acoustics

Series Aalto University publication series DOCTORAL DISSERTATIONS 144/2011

Field of research Acoustics and Audio Signal Processing

Manuscript submitted 30 May 2011 **Manuscript revised** 15 November 2011

Date of the defence 20 January 2012 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

The periodicity of speech sounds which is produced by the vibration of the vocal folds, plays a significant role in speech communication. In the auditory system, sound periodicity is extracted along the neural pathway and is, according to several studies of the human brain, represented in the cortical level by a periodicity-specific neural population. Such a population could encode the periodicity of speech sounds. The evidence for cortical periodicity-sensitivity, however, rests mostly on measures of brain activity elicited by non-speech stimuli that differ from speech sounds with respect to their acoustic features and perceptual qualities. Thus, the generalizability of these results to natural speech communication may be limited.

The work presented in this thesis investigated cortical processing of the periodicity of speech sounds by using controlled manipulations in the periodicity of vowel stimuli and by measuring brain activity elicited by these stimuli with magnetoencephalography. The results indicate larger amplitudes and more anterior source locations for the responses elicited by periodic as opposed to aperiodic vowel stimuli. While such an effect of periodicity was observed for a range of fundamental frequencies (F0), degrees of periodicity, and durations of the periodic vowel stimuli, the cortical periodicity-specific activity was also modulated by these parameters. Furthermore, evidence for aperiodicity-sensitive activity was found through stimulus-specific release from adaptation when aperiodic vowel stimuli were alternated with periodic rather than with aperiodic adaptors.

The results of the thesis, thus, indicate that the degree of speech sound periodicity, determined by the vocal fold vibration, is represented in the auditory cortex. Such sensitivity to periodicity might reflect the activity of distinct neural populations that are selective to sound periodicity and aperiodicity. Importantly, this view of distinct feature-selective populations can, based on the current results, be generalized to describe the neural mechanisms of speech perception. The dependency of the observed periodicity-sensitivity on the acoustic features of the vowel stimuli, further, appears to reflect cortical encoding of auditory-perceptual aspects of voice quality.

Keywords periodicity, pitch, magnetoencephalography, N1m, sustained field

ISBN (printed) 978-952-60-4443-9 **ISBN (pdf)** 978-952-60-4444-6

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2011

Pages 138 **The dissertation can be read at** http://lib.tkk.fi/Diss/

Tekijä
Santeri Yrttiaho

Väitöskirjan nimi
Puheäänten jaksollisuuden käsittely kuuloaivokuorella

Julkaisija Sähkötekniikan korkeakoulu

Yksikkö Signaalinkäsittelyn ja akustiikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 144/2011

Tutkimusala Akustiikka ja äänenkäsittelytekniikka

Käsikirjoituksen pvm 30.05.2011 **Korjatun käsikirjoituksen pvm** 15.11.2011

Väitöspäivä 20.01.2012 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Puheäänten jaksollisuudella on tärkeä merkitys puhekommunikaatiossa. Kuulojärjestelmässä äänen jaksollisuus tunnistetaan kuuloradan eri vaiheissa tapahtuvan laskennan avulla. Lisäksi aikaisemman aivotutkimuksen mukaan kuuloaivokuorella sijaitsee todennäköisesti äänen jaksollisuudelle herkkä solupopulaatio, joka voisi osallistua myös puheäänten jaksollisuuden käsittelyyn. Kyseiset tutkimustulokset kuitenkin liittyvät valtaosin muiden kuin puheäänten hermostolliseen käsittelyyn. Kuitenkin, koska puheäänet poikkeavat muista ääniärsykkeistä sekä akustisten piirteiden että kuulohavainnon osalta, edellyttää luonnolliseen puhekommunikaatioon yleistettävä kuvaus puheen havaitsemisen aivomekanismeista puheärsykkeiden käyttöä.

Tämän väitöskirjan tutkimukset keskittyivät puheen jaksollisuuden käsittelyyn kuuloaivokuorella. Näissä tutkimuksissa vokaaliärsykkeiden jaksollisuutta muokattiin soveltamalla puheen akustista mallinnusta ja näihin ärsykkeisiin liittyvää aivotoimintaa mitattiin magnetoenkefalografialla. Tulokset osoittivat jaksollisten vokaalien tuottavan aivotoimintaa, joka on voimakkaampaa ja paikantuu edemmäksi kuin epäjaksollisiin vokaaliärsykkeisiin liittyvä aivotoiminta. Vaikka vastaava herkkyys jaksollisuudelle säilyi useilla vokaaliärsykkeiden perustaajuuksilla, äänenkestoilla ja jaksollisuuden asteilla, kuuloaivokuoren jaksollisuudelle herkän toiminnan todettiin kuitenkin riippuvan yllämainituista akustisista piirteistä. Tutkimalla peräkkäisten ääniärsykkeiden aiheuttamien aivovasteiden välistä vuorovaikutusta, havaittiin edelleen paitsi puheen jaksollisuudelle myös sen epäjaksollisuudelle herkkää aivotoimintaa.

Väitöskirjan tulokset osoittavat kuuloaivokuoren toiminnan heijastavan puheäänten jaksollisuutta. Näiden tulosten mukaan teoria, jonka mukaan kuuloaivokuorella sijaitsee erillisiä äänen jaksollisuudelle ja epäjaksollisuudelle herkkiä hermopopulaatioita voisi olla yleistettävissä koskemaan myös puheäänten käsittelyä. Edelleen kuuloaivokuoren jaksollisuudelle herkän aivotoiminnan riippuvuus vokaaliärsykkeiden piirteistä näyttäisi heijastavan puheen äänenlaatua koskevan havainnon muodostumista.

Avainsanat jaksollisuus, äänenkorkeus, magnetoenkefalografia, aivovasteet

ISBN (painettu) 978-952-60-4443-9 **ISBN (pdf)** 978-952-60-4444-6

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2011

Sivumäärä 138 **Luettavissa verkossa osoitteessa** http://lib.tkk.fi/Diss/

Acknowledgements

The research for this thesis was conducted at the Department of Signal Processing and Acoustics at Aalto University School of Electrical Engineering in collaboration with the BioMag Laboratory at Helsinki University Central Hospital and the Department of Biomedical Engineering and Computational Science (BECS) at Aalto University School of Science. The research was funded by the national Graduate School in Electronics, Telecommunication and Automation (GETA), the Academy of Finland, and the Emil Aaltonen Foundation.

I wish to thank the preliminary examiners, Professors Matti Hämäläinen and Eric Schröger, for their reviews which helped me to improve my thesis and my thinking about brain research.

I express my gratitude for my supervisor, Professor Paavo Alku, for inviting me into the fascinating world of speech science and for providing me the exceptional opportunity to work as a part of a multidisciplinary research collaboration in the field of auditory neuroscience of speech communication. His example and his ability to build confidence in his students have been crucial for the completion of my thesis. In this collaboration, what I have learned from Docents Patrick May and Hannu Tiitinen has been enormous. As instructors of my Ph.D. research they have provided me with a great deal of support, encouragement, and vision for which I am sincerely grateful.

I thank Docent Jyrki Mäkelä for welcoming me into his underground lair, the BioMag laboratory, where all the magnetoencephalography data for my thesis was collected. The warm atmosphere and the help received from the staff, especially Juha Montonen, made the time spent in BioMag cheerful and productive.

My other “home” at the Aalto University besides the “Acoustics Lab” which, by the way, seems to be attracting many brilliant and inspiring people like Hannu Pulakka, Jouni Pohjolainen, Emma Jokinen, Tuomo Raitio, and Dr Daniel Aalto, was BECS. The work of Professors Mikko Sams and Risto Ilmoniemi at BECS has provided outstanding settings for cutting-edge brain research. I also wish to thank Professor Sams for the encouragement and for inviting our group to join the Mind and Brain Laboratory (MBL). A big thank you also goes for all the MBL members.

During my Ph.D. work I received influential Peer review, Peer support, and positive Peer pressure, if you like, from Dr Nelli Salminen and Ismo Miettinen. The years when we shared the moments of joy, disappointment, and hopes for the future when learning to become scientists will never be forgotten.

I am also grateful to my parents and to my brother for being able to grow up in an environment from which my orientation to scientific and academic work emerges naturally. Finally, my warmest gratitude goes to my wife, Kirsi, for all the love and support and to my daughter, Isabella, for making every time daddy comes home from work a happy time.

Espoo, December 2011

Santeri Yrttiaho

Contents

List of publications

This thesis is based on the following publications:

I Yrttiaho, S., Tiitinen, H., May, P. J. C., Leino, S., and Alku, P. (2008). Cortical sensitivity to periodicity of speech sounds. *Journal of the Acoustical Society of America*, 123:2191-2199.

II Yrttiaho, S., Alku, P., May, P. J. C., and Tiitinen, H. (2009). Representation of the vocal roughness of aperiodic speech sounds in the auditory cortex. *Journal of the Acoustical Society of America*, 125:3177-3185.

III Yrttiaho, S., Tiitinen, H., Alku, P., Miettinen, I., and May, P. J. C. (2010). Temporal integration of vowel periodicity in the auditory cortex. *Journal of the Acoustical Society of America*, 128:224-234.

IV Yrttiaho, S., May, P. J. C., Tiitinen, H., and Alku, P. (2011). Cortical encoding of aperiodic and periodic speech sounds: evidence for distinct neural populations. *Neuroimage*, 55:1252-1259.

Author's contributions

Study I The candidate was the first author of the publication and was responsible for the collection and the analysis of data from the magnetoencephalography (MEG) and the behavioral experiment. The behavioral experiment was designed by the candidate. The study plan and the design of the MEG experiment were delineated by the supervisor, professor Paavo Alku, and the instructors of the thesis, docent Hannu Tiitinen and docent Patrick May. All contributing authors provided valuable feedback for the candidate in writing the article.

Study II The planning of the study and the design of the MEG and the behavioral experiments were done by the candidate. The candidate was also the first author of the publication and collected as well as analyzed all of the experimental data. Professor Alku, and docents Tiitinen and May provided valuable instruction throughout the investigations and in writing the article.

Study III The study plan and the design of the experiments were done by the candidate. He was the first author of the publication and collected as well as analyzed all of the experimental data. The auditory model included in the publication was also designed by the candidate. Professor Alku, and docents Tiitinen and May provided valuable feedback throughout the investigations. All contributing authors provided important insights and feedback for the candidate in writing the article.

Study IV The study plan and the design of the MEG experiment were due to the candidate. The principle authorship of the publication was that of the candidate who also collected and analyzed the experimental data. Professor Alku, and docents Tiitinen and May provided valuable feedback in writing the article.

List of abbreviations

ACF	Autocorrelation function
AEF	Auditory evoked field
BOLD	Blood oxygen level dependent
CAPE-V	Consensus auditory-perceptual evaluation of voice
CN	Cochlear nucleus
ECD	Equivalent current dipole
EEG	Electroencephalography
EGG	Electroglottography
F0	Fundamental frequency
fMRI	Functional magnetic resonance imaging
GIF	Glottal inverse filtering
GRBAS	Grade, roughness, breathiness, asthenia, and strain
HG	Heschl's gyrus
IC	Inferior colliculus
IRN	Iterated ripple noise
ISG	Interstimulus gap
ISI	Interstimulus interval
lHG	Lateral Heschl's gyrus
MEG	Magnetoencephalography
N1	A vertex-negative auditory evoked potential at around 100 ms latency
N1m	The magnetic counterpart of N1
OQ	Open quotient
PET	Positron emission tomography
PP	Planum polare
PSE	Point of subjective equality
PT	Planum temporale

RFN	Recycling frozen noise
SF	Sustained field
SP	Sustained potential
SQ	Speed quotient
SQUID	Superconducting quantum interference device
SSA	Stimulus-specific adaptation
STG	Superior temporal gyrus
T	Period length
t_c	Length of the glottal closed phase
t_{o1}	Length of the glottal opening phase
t_{o2}	Length of the glottal closing phase
TWI	Temporal window of integration

1. Introduction

Speech communication is based on the ability of the human voice production mechanism to produce acoustically diverse utterances and on the ability of the auditory system to encode this diversity. The features of speech sounds derive from two separate time-variant processes which are called phonation and articulation (Fant, 1960). The linguistic categorization of speech sounds is largely based on the resonant frequencies, or formants, of speech that are determined by the vocal tract during articulation. A fundamental aspect in phonation, in turn, is the function of the vocal folds which may vibrate at different frequencies and with differing degrees of periodicity during the production of voiced speech sounds such as vowels. The vocal fold vibration can also be completely absent during the production of aperiodic speech sounds such as fricative consonants and during whispering. The variability of the vocal fold vibration enables many significant features in speech communication including melodic contours or intonation (Hirst and Di Cristo, 1998), signaling the gender, age, size, and emotional state of the speaker (Murray and Arnott, 1993; Titze, 1994), tonal languages (Yip, 2002), and the phonemic contrast between voiced and unvoiced speech sounds (Pickett, 1999). Therefore, the periodicity of speech sounds, produced by the vibration of the vocal folds, is an important aspect of speech communication.

The perception of sound periodicity has been the subject of an enduring scientific investigation for at least since the controversy between Seeback and Ohm in the 19th century regarding the pitch perception of the missing fundamental frequency (de Cheveigné, 2005). From this background, the theories of periodicity processing have been intertwined with those regarding elementary aspects of the auditory nervous system such as frequency selectivity or tonotopy (Winter, 2005) and phase locking to the sound waveform (Plack and Oxenham, 2005a). On the one hand, according to the tonotopic account, the perception of the pitch of periodic sounds depends on the spatial pattern of displacement of the basilar membrane in the cochlea. The proponents of the so-called temporal models of pitch, on the other hand,

argue that pitch perception arises from the ability of auditory neurons to phase-lock, that is, to synchronize their discharge patterns to the time-domain features of the stimulus. Despite the efforts in auditory research, the neural mechanisms underlying the perception of sound periodicity have remained controversial. In particular, the relative importance of phase-locked (de Cheveigné, 2005) versus tonotopic (Shamma, 2001) representations in the neural encoding of periodicity remains a debated topic.

There are several acoustic features that affect the perceived pitch of a periodic sound (Fastl and Zwicker, 2007). These include the degree of periodicity (similarity between successive cycles of the stimulus), the fundamental frequency (F0; which is the inverse of the cycle length of the stimulus), and the order of the harmonics (integer multiples of the F0) that are present in the sound spectrum. Many of these features can be used to characterize any audible sound and, thus, the investigation of the perceptual encoding of these features comes with the challenge of choosing a representative auditory stimulus. Traditionally in auditory research, the emphasis has been in strict experimental control over the key acoustic variables of the stimuli. Consequently, most of the research has been conducted with artificial non-speech stimuli such as click trains and pure tones which are never encountered in natural auditory environments (Neuhoff, 2004).

Already in the 1960s, Liberman (1967) proposed that the perception of speech is based on dedicated mechanisms that are different from those used in generic auditory perception. In this light, the results from experimental studies with synthetic non-speech stimuli might not apply to the perception of speech sounds. Although the view of speech specificity in the scope of Liberman's motor theory is controversial (*e.g.*, Massaro and Chen, 2008), recent brain imaging shows cortical activation that is specific to speech (Whalen *et al.*, 2006) or to human vocalizations *per se* (Belin *et al.*, 2000). The perception of sound periodicity has, further, been shown to depend on the phonemic identity or the formant structure of vowel stimuli (Chuan and Wang, 1978; Hellström *et al.*, 1994; Robinson and Patterson, 1995). For the brain research of speech perception these results, thus, call for the use of representative stimulus material that matches the acoustic and perceptual qualities of natural speech.

The investigation of periodicity-sensitive activity in the brain requires defining the criteria for determining the observed activity as periodicity-sensitive. The most reasonable criterion seems to be the dependency of the neural activity on the F0 or on the degree of sound periodicity (Bendor and Wang, 2010). The activity of a periodicity-sensitive neuron tuned to a specific F0 increases when the F0 of a stimulus coincides with the frequency preferred

by the neuron. Thus, sensitivity to periodicity could be identified through F0 tuning curves. However, the dependency of neural activity on the stimulus F0 can be related to other, confounding, features of the stimulus such as the locus of the maximum spectral peak(s). For example, in pure tone stimuli the locus of the spectral maximum correlates perfectly with the F0 of the stimulus. Hence, the variability of neural activity elicited by such stimuli might indicate purely tuning of the neural activity to a certain frequency region irrespective of the periodicity or the F0 of the stimulus *per se*. Furthermore, Bendor and Wang (2010) have suggested that a large proportion of the modulation-sensitive neurons found throughout the auditory cortex encode the average repetition rate of complex sounds but are generally insensitive to the degree of sound periodicity. In this light, sensitivity to the degree of periodicity constitutes a critical determinant of periodicity-sensitive neural activity, and may in some cases serve as an even more adept criterion for periodicity-specific activity than the sensitivity to the F0 of a stimulus. Further, in addition to the F0, the degree of periodicity is a communicatively significant feature of speech. Thus, identifying the neural mechanisms that are sensitive to the degree of sound periodicity is essential for understanding the cerebral encoding of the features of speech sounds that are related to periodicity.

The aim of this thesis was to study cortical representations of speech periodicity. The auditory cortical activity was measured non-invasively in human subjects with magnetoencephalography (MEG) which registers the magnetic fields produced by electric currents in the brain. The auditory stimuli consisted of vowel sounds that are representative of the sounds present in natural speech communication with respect to their relevant acoustic and perceptual qualities.

2. Periodicity of speech sounds

2.1 The origin of speech periodicity in speech production

Speech sounds can be distinguished based on the manner and the place of articulation as well as on the presence of the voicing feature (Pickett, 1999). Flanagan (1972) classified speech sounds into three main categories by using such labels. These categories include voiced, unvoiced, and plosive sounds. The voicing feature is determined by the involvement of the periodic vibration of the vocal folds in speech production which is present and absent in the voiced and unvoiced sounds, respectively (Ladefoged and Maddieson, 1996). In contrast to sustained voiced (such as vowels, *e.g.*, [a]) or unvoiced sounds (such as unvoiced fricatives, *e.g.*, [s]), the plosives (*e.g.*, [p]) are produced by an abrupt release of air blocked by the vocal tract. Plosives, or stop consonants, may be further categorized into voiced plosives, such as [b], and into unvoiced plosives, such as [p], depending on the posture of the vocal folds during the closed phase of the consonant (Pickett, 1999). These physiological factors related to speech production are correlated with distinct acoustic characteristics that distinguish between voiced, unvoiced, and plosive speech sounds. Importantly, the voiced speech sounds, unlike the unvoiced sounds, are characterized in the time-domain by a periodic structure. Such periodic sounds are prevalent in spoken language and constitute around 78 % of spoken English (Catford, 1977). The significance of the voiced, periodic speech sounds is accentuated by the longer duration and higher energy of these sounds in comparison to the other categories of speech sounds. Moreover, the voiced sounds, especially vowels, have an important phonetic role in most languages. Due to coarticulation, the voiced sounds also contain phonemic information about the surrounding voiced, unvoiced, and plosive sounds by means of transitions in the formant frequencies (*e.g.*, Öhman, 1966).

In speech production, the airflow from the lungs is passed through the vocal folds into the vocal tract which includes the pharynx, mouth, teeth, tongue, and lips (Fig. 1). The aspects of speech production that are determined by the function of the vocal folds can be described as phonation in contrast to the

process of articulation where the contribution of the vocal tract is dominant (Fant, 1960). The vocal folds, or thyro-arytenoid ligaments, are located in the larynx and stretch between the arytenoid cartilage and the thyroid cartilage (referred to as the "Adam's apple" in adult males). The paired structure of vocal folds is divided by the glottis, which is an aperture located between the vocal folds that can be opened or closed depending on the configuration of the arytenoid cartilages and on the subglottal pressure. The periodicity of voiced speech sounds originates from the periodic vibration of the vocal folds which is maintained by the pulmonary airflow (Titze, 1994; Pickett, 1999). In this vibration, the vocal folds rapidly and periodically contact each other and then separate, thus producing the opening and closing of the glottis. The vocal fold vibration is self-sustained due to the coupling of the aerodynamics of the trachea and the vocal tract. That is, the inertia of the air accumulating in the vocal tract provides positive feedback for the vocal fold vibration (Titze, 1994). As a result of the vocal fold vibration, the airflow passing through the glottis is modulated so that the flow increases in the opening phase and then decreases promptly in the glottal closing phase. The airflow signal provided by the glottal pulsation into the vocal tract is called the glottal excitation or the glottal volume velocity waveform. This signal acts as the raw material for the articulation of vowel sounds and gives the general spectral shape that constrains the spectra of the vowel sounds (Pickett, 1999). The periodic glottal excitation can be described in the time domain as having smooth pulses characterized with a more or less negative skew (Fig. 2), and in the frequency domain as having decreased power towards higher frequencies (Fig. 1, bottom right) with a negative spectral tilt in the range of around 6–18 dB/octave (Childers and Lee, 1991).

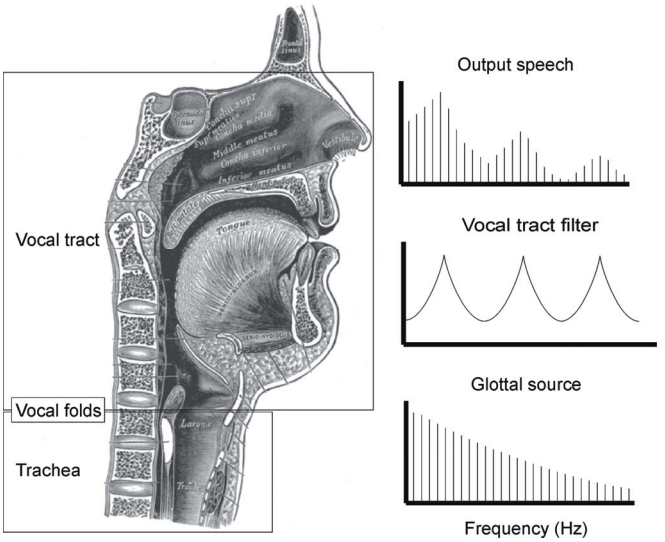

Figure 1. The main anatomic structures of the human voice production system (Adapted from Gray and Lewis, 1918) along with a schematic illustration of the source-filter model of voice production. Sagittal section of the trachea, vocal folds, and vocal tract is shown on the left. The spectra of the glottal excitation, vocal tract filter, and output speech according to the source filter model are shown on the right.

The vibration of the vocal folds is not perfectly periodic even during the production of voiced speech sounds. These sounds can, thus, be more accurately described as being quasiperiodic (Rabiner and Schafer, 2007). The deviations in voiced speech from perfect periodicity are often increased in vocal fold pathologies (Lieberman, 1963; Iwata and von Leden, 1970; Murry and Doherty, 1980) or in creaky voices produced in the so-called vocal fry register (Henton and Bladon, 1988). In addition, a special case of aperiodic phonation occurs in whispering where the vocal folds are simultaneously being adducted and prevented from vibrating which produces turbulent high-velocity airflow into the pharynx (Laver, 1994).

The length of the vocal folds depends on the gender of the speaker, it being longer in males (~1.6 cm, Hirano *et al.*, 1981; 1.75–2.50 cm, Titze, 1994) than in females (~1 cm, Hirano *et al.*, 1981; 1.25–1.75 cm, Titze, 1994). The vocal folds are typically 0.3–0.5 cm thick (Hahn *et al.*, 2006) and have a mass of around 1 gram (Hirano *et al.*, 1981). These factors determine the range of F0s that a given speaker is capable of producing as the average F0 of speech decreases with increasing length and mass of the speaker's vocal folds (Welham, 2009). In order to produce speech sounds with variable F0s, a speaker can adjust the function of the vocal folds with laryngeal muscles that

control the length and the stiffness of the vibrating tissue of the vocal folds (Titze, 1994). The F0 of voiced speech sounds increases mainly with increasing tension exerted on the vocal folds which may be produced by constricting the small muscles of the larynx. The F0 is also influenced by the subglottal pressure and by the distribution of the vocal fold mass (Pickett, 1999). By adjusting these parameters, a speaker can vary the frequency of the vocal fold vibration over broad ranges which center around 125 Hz, 200 Hz, and 500 Hz in male, female, and infant speakers, respectively (Titze, 1994).

The voiced speech sounds can be characterized by different voice types such as modal, vocal fry, pressed, breathy, and falsetto (Childers and Lee, 1991, Alku and Vilkman, 1996). From the physiological perspective, these characteristics are related to the positioning of the arytenoid cartilages (Pickett, 1999) which affects the length and the thickness of the vocal folds (Childers and Lee, 1991). Increased rotation and adduction of the cartilages results in a so-called pressed voice while a breathy voice is produced by a lax adduction of the arytenoids (Pickett, 1999). The falsetto, in turn, is produced by increasing the length and by simultaneously decreasing the thickness of the vocal folds (Childers and Lee, 1991) so that the vibrating mass and the vibratory amplitude are reduced (Henrich, 2006). These voice types have distinct patterns of vocal fold vibration (Fig. 2) that can be described with factors such as the relative durations of the glottal opening and closing phase, abruptness of the glottal closure, and in terms of the spectral tilt of the glottal excitation (Childers and Lee, 1991; Alku and Vilkman, 1996).

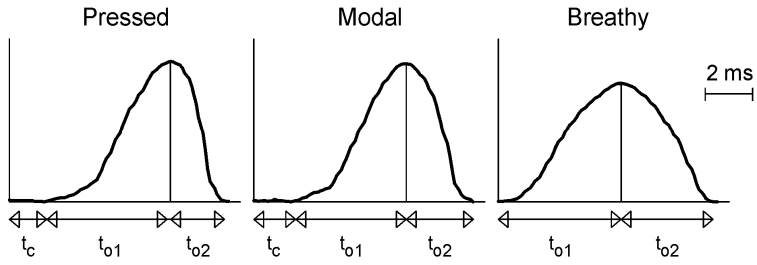

Figure 2. Glottal pulses in pressed, modal, and breathy phonation. The three types of phonation are distinguished by the duration of the glottal closed phase (t_c), opening phase (t_{o1}), and closing phase (t_{o2}) relative to the period length. The duration of especially the closing phase decreases from breathy to modal and from modal to pressed phonation. Adapted from Alku and Vilkman (1996).

The modal voice, resulting from normal phonation, is characterized by medium values of the glottal parameters, and the other voice types can be

viewed as deviations from the modal voice. The pressed voice, produced by tensing the vocal folds, yields an excitation with short glottal pulses characterized spectrally with increased harmonic richness (Alku and Vilkman, 1996). The glottal pulses in breathy and falsetto voices are broadened and the spectral tilt is increased relative to the modal voice (Childers and Lee, 1991). The breathy phonation may also involve high frequency noise and amplitude perturbations (Childers and Lee, 1991), that is, random variations in the pulse amplitude.

Voice type is partly independent of the F0 of speech. In particular, in the pressed–modal–breathy continuum, voices can be produced without changing the F0 (Alku and Wilkman, 1996) and only a weak correlation between an amplitude-domain index of phonation type (Alku *et al.*, 2002) and F0 has been observed (Campbell and Mokhtari, 2003). The glottal excitation pattern, however, tends to vary with voice registers that the speakers use to produce speech in different F0 regions. Speech with very low F0 (18–52 Hz) is typically produced in the vocal fry register with drawing the arytenoid cartilages closely together (Henton and Bladon, 1988). While each glottal cycle in modal phonation contains a single pulse, in vocal fry, an individual cycle may have one to three separate glottal pulses (Whitehead *et al.*, 1984; Childers and Lee, 1991). Furthermore, frequency perturbations, that is, random variations in the period length of the glottal excitation are common in vocal fry (Henton and Bladon, 1988).

The variable sounds produced by the vibration of the vocal folds are strongly filtered by the vocal tract in the frequency domain. The shape of the vocal tract determines its resonant frequencies, that is, the formants which the speaker can adjust by varying the openness of the oral and the nasal cavities and by moving the tongue into different positions (Ladefoged and Maddieson, 1996; Pickett, 1999). This process of articulation is used to produce phonetically distinct speech sounds that can be identified by a listener based on the values of the formant frequencies. According to the source-filter model of speech production (Fant, 1960), the glottis (source) and the vocal tract (filter) function in a mutually independent fashion and constitute a linear process. Although this assertion does not entirely hold, the quasiperiodic signal produced by the glottal pulsation can be investigated largely independently of the vocal tract function with inverse filtering applications of the source-filter theory (*e.g.*, Alku *et al.*, 1999).

Glottal inverse filtering (GIF) refers to various techniques that can be employed to extract the glottal excitation from the speech signal. Due to the concealed location of the vocal folds and the filtering effect of the vocal tract, the glottal pulsation cannot normally be observed directly either visually or acoustically. Thus, GIF involves an inverse problem where the input to the

vocal tract (glottal excitation) is to be estimated when only the output from the vocal tract (speech) is known. According to the source-filter model of speech production, this problem of estimating the source from the output can be solved by cancelling out the effects of the filter. That is, when the effects of the vocal tract filter has been removed from the output speech, the signal corresponding to the glottal excitation remains. Therefore, GIF requires first extracting the effects of the vocal tract from the speech signal and then cancelling out these effects of the vocal tract filtering from the speech waveform. While GIF can be facilitated with instruments such as electroglottography (EGG; Childers *et al.*, 1990), the glottal excitation can be estimated non-invasively from purely acoustical analyses of the speech waveform as well (Strube, 1974; Alku, 1992; Fröhlich *et al.*, 2001; Alku *et al.*, 2009).

2.2 The relationship between speech periodicity and voice quality

Laver (1980) defines voice quality as "the characteristic auditory colouring of a speaker's voice". Voice quality refers to those features of speech that do not convey linguistic information but which instead signal the identity, personality, age, health, and other physical characteristics of the speaker (Story *et al.*, 2001). These features are influenced by virtually all of the processes that take place during speech production in the respiratory, phonatory, and in the articulatory systems. Consequently, voice quality is a multidimensional aspect of speech communication. Importantly, the laryngeal aspects of vocal quality, related to the characteristics of the vibratory pattern of the vocal folds, exert a major influence on the quality of speech (Laver, 1980).

The laryngeal voice quality can be studied objectively with acoustic, EGG, aerodynamic, and imaging analyses (Childers and Lee, 1991; Blomgren *et al.*, 1998; Airas, 2008). An important feature in laryngeal voice quality is the degree of periodicity in the vibratory pattern of the vocal folds which determines the degree of periodicity of the output speech. Typically, the degree of sound periodicity has been studied with acoustic analyses of the speech waveform without inverse filtering. Such acoustic analyses confirm the notion that the voiced sounds produced by healthy speakers are characterized by a high degree of periodicity. That is, only small amounts of random variability is present in the speech waveform, and correspondingly, in the glottal pulsation during voicing (*e.g.*, Horii, 1979; Blomgren *et al.*, 1998; Muñoz *et al.*, 2003). The random variability in the glottal excitation, reflected

in the speech waveform, may, however, be increased in non-modal voice production (Childers and Lee, 1991; Laver, 1994; Blomgren *et al.*, 1998), in laryngeal pathologies (Lieberman, 1963; Iwata and von Leden, 1970; Murry and Doherty, 1980), or due to aging (Vipperla *et al.*, 2010). Such aperiodicities can be categorized as random variations in the intervals (jitter) or amplitudes (shimmer) of successive glottal pulses or by additive noise (harmonics-to-noise ratio) (Baken and Orlikoff, 2000).

The complete understanding of voice quality requires both the objective metrics that characterize voices and the "auditory colourings" perceived by the listener. To describe the perception of speech aperiodicites, terms such as roughness, harshness, hoarseness and breathiness have been used. To some extent this vocabulary can be related to specific phonatory or acoustical characteristics. For example, roughness has typically been used to describe voices with jitter or shimmer while breathiness describes the perceptual quality of a voice with added noise due to audible air escape (*e.g.*, Dejonckere, 2010). However, the terminology on disordered voice quality has not always been used consistently (Kempster *et al.*, 2009). The mapping between acoustic characteristics and perceptual quality is further complicated by the significant intercorrelations of voice aperiodicities (Horii, 1980) and the difficulties in segregating the aperiodicities in acoustic analyses (Hillenbrand, 1987; Michaelis *et al.*, 1998). When studying the perceptual aspects of voice quality, these difficulties can be avoided by using synthetic stimuli in which speech aperiodicities can be varied independently. The results from such studies indicate strong relationships between stimulus aperiodicities and perceived degradation in quality (Wendahl, 1966; Hillenbrand, 1988). The contribution of individual types of voice aperiodicity to voice quality may, however, change in the presence of other aperiodicities (Kreiman and Gerrat, 2005). Recognizing these difficulties, researchers have attempted to provide a consistent protocol for the analysis of voice quality to be used in clinical and scientific investigations of auditory-perceptual aspects of voice quality (Hirano, 1981; Kempster *et al.*, 2009; Dejonckere, 2010). A recent assessment tool, Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V), identifies six main perceptual dimensions of voice quality: overall severity, roughness, breathiness, strain, pitch, and loudness (Kempster *et al.*, 2009). The definitions provided for these dimensions in CAPE-V are shown in Table I.

Table I. Perceptual dimensions of voice quality in CAPE-V (Kempster *et al.*, 2009).

Overall Severity:	Global, integrated impression of voice deviance
Roughness:	Perceived irregularity in the voicing source
Breathiness:	Audible air escape in the voice
Strain:	Perception of excessive vocal effort (hyperfunction)
Pitch:	Perceptual correlate of fundamental frequency
Loudness:	Perceptual correlate of sound intensity

Besides deciding on the perceptual dimensions (such as roughness and breathiness) that are assessed by the listeners, the investigation of auditory-perceptual aspects of voice quality requires specifying the variables related to the assessment process *per se* (Kempster *et al.*, 2009). This includes choosing the type of listeners involved and the training that is required from the listeners, choosing the appropriate type of speech material to be assessed (vowels, words, sentences, or spontaneous speech), and the type of task that the listeners perform to make the evaluations. While these variables allow for many variations in the design of voice assessments, there are standard processes of evaluation of voice quality for specific application areas such as medical voice problems (Kempster *et al.*, 2009) and telecommunications (ITU-T recommendation P.800, 1996). Voice quality can be assessed by using a numerical rating scale (Hillenbrand, 1988; Kreiman *et al.*, 2007), a visual analog rating (Kreiman *et al.*, 2007; Kempster, 2009), paired comparisons (Wendahl, 1966; ITU-T recommendation P.800, 1996), or adjustment procedures (Kreiman and Gerrat, 2005). An important feature in the assessment protocol is whether the listeners are provided with comparison stimuli or not. Kreiman and colleagues (2007) investigated systematically the effects of this aspect of task design on the reliability of voice quality evaluations. In their study, listeners assessed vocal quality in tasks that differed in the presence/absence of comparison stimuli and in the extent to which the comparison stimuli (if present) matched the target voices. They observed that the most accurate ratings of voice quality were obtained from conditions with reference stimuli that were closely matched to the target voices. Thus, their results suggest that a large portion of interrater variability in voice quality judgments is due to the task design rather than to the unreliability of the listeners. Therefore, with a suitable procedure, reliable

ratings of perceptual voice quality can be obtained from listeners with minimal training.

The laryngeal voice quality can be described not only in terms of the presence or absence of aperiodicities but also in terms of voice types (*e.g.*, modal, pressed, vocal fry, breathy, and falsetto). The physiological differences in producing these voice types and their acoustic correlates were described in Section 2.1. The acoustical differences between voice types are also perceptually relevant. This is evident, for example, from the inclusion of a dimension called strain in the labeling systems of voice quality such as CAPE-V (Kempster *et al.*, 2009) and GRBAS (Hirano, 1981) where the term is used to refer to a pressed voice produced with excessive vocal effort. Moreover, in a study of modal voice and vocal fry, listeners (N=55) were able to identify the two voice types with 95.5 % and 100 % accuracy, respectively, from sustained vowels [a] (Blomgren *et al.*, 1998). Finally, Childers and Lee (1991) showed that the perception of the degree of vocal strain (tense vs. lax) and breathiness was correlated with the parameters of the glottal waveshape and added noise of synthesized vowels [a], respectively.

In summary, the laryngeal component in speech production, which constitutes the origin of the periodicity of voiced speech sounds, has an important role in establishing the vocal quality of speech. The phenomenon of voice quality is multidimensional and must be approached from the perspectives of speech physiology, acoustics, and perception. The relationships between different dimensions of voice quality (*e.g.*, jitter, shimmer, and voice type) and between physiological, acoustical, and perceptual aspects of voice quality are, furthermore, often complicated. Nevertheless, the vibratory pattern of the vocal folds and the aperiodicities in speech waveform that contribute to voice quality can be quantified with using objective metrics that are related to perceptual voice quality.

2.3 Linguistic aspects of speech periodicity

The phonetic features related to the laryngeal aspects of production, acoustics, and perception of speech sounds can be mapped onto phonological categories that are used to encode meaning in spoken language (Bloch, 1941; Pickett, 1999). In linguistics, speech has been hierarchically divided into units of different sizes (*e.g.*, phonemes, syllables, words) depending on the amount of lower level units that they contain. The most fundamental linguistic unit in such a hierarchy is the phoneme, which is the smallest unit that can be used to form linguistic contrasts (*e.g.*, Bloomfield, 1933; Pickett, 1999). For example,

the words "*me*" and "*we*" are differentiated only by one phoneme and, thus, constitute a minimal pair. Phonemes, in turn, can be distinguished on the basis of their voicing so that voiced sounds produced by vibrating the vocal folds periodically are contrasted to unvoiced sounds produced without this vibration. This contrast separates, for example, the consonant /s/ from /z/ as the production of only the latter involves vocal fold vibration. Furthermore, the presence or absence of periodicity can constitute an additional divergent feature between pairs of phonemes which already differ with respect to other features. For example, all vowel sounds differ form the unvoiced consonants in that the former are produced predominately with and the latter without vibrating the vocal folds periodically.

In addition to the voicing contrast, the variability of the F0 of the periodic speech sounds can also play a linguistic role. In tonal languages (*e.g.*, Mandarin, Thai, Yoruba, and Swedish) individual words can be differentiated based on their F0 or their F0 patterns (Yip, 2002). While tonal features are used in half of the world's languages (Felder *et al.*, 2009) their importance and structure vary between languages. For example, while the phonology of Mandarin Chinese relies heavily on five distinct tones, Standard Swedish uses two tones (a low or late peaking *accent I* and a high or double peaking *accent II*; Schaeffler, 2005) that separate around 350 minimal pairs (Elert, 1972). In Mandarin the use of tonal contrasts more than triples the amount of syllabic minimal pairs (Duanmu, 2007) so that thousands of word pairs can be distinguished based on tonal contrasts. In many languages where the segmental level (identification of the vowels and consonants) is sufficient to determine the correct word, tonal features are used as additional information (Felder *et al.*, 2009; Vainio *et al.*, 2010) and the division between tonal and non-tonal languages may be viewed as a continuum rather than as an absolute one (Yip, 2002). Finally, intonational variability in the F0 during the production of complete sentences is, further, used to differentiate between pragmatic categories such as statements, questions, and orders (Hirst and Di Cristo, 1998; Patel, 1998) independently of tonality.

3. Periodicity and pitch perception

3.1 Psychoacoustics of pitch perception

The waveform of periodic speech sounds such as vowels is characterized by a periodic oscillation. In hearing, this oscillation fuses into a steady perception of a continuous, definite, pitch. According to the American National Standards Institute (ANSI) definition: “Pitch [is] that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low. Pitch depends primarily on the frequency content of the sound stimulus --” (ANSI, 1994; in Plack and Oxenham, 2005b).

The lower and upper limits of pitch perception have been studied with tasks where the subject is either directly asked to report the audibility of the pitch sensation (*e.g.*, Ritsma, 1962) or with F0 discrimination tasks (*e.g.*, Krumbholz *et al.*, 2000). Pitch salience and the ability to detect changes in F0 seem to be related phenomena as both measures converge into similar estimates of the lower limit of pitch (Krumbholz *et al.*, 2000). While the lower (F0) limit of pitch perception depends on the sound spectrum, in favorable conditions it is located in the 20–40 Hz frequency range (40 Hz, Ritsma, 1962; 19 Hz Guttman and Julesz, 1963; 30 Hz, Krumbholz *et al.*, 2000; 30 Hz, Pressnitzer *et al.*, 2001). The lowest, 19-Hz estimate for the pitch limit by Guttman and Julesz (1963) was obtained with recycling frozen noise (RFN) stimuli, which are constructed by concatenating identical segments of wideband noise with segment duration of 1/F0. For very high F0s, that is, above 5000 Hz, the perception of pitch is also affected or disappears altogether (Ward, 1954; Semal and Demany, 1990).

For sounds with very low F0 values relative to those typically used in speech communication (see Section 2.1), the individual cycles of the sound become audible and the perception of pitch disappears. In particular, vocal fry speech with very low F0s, typically yields a creaky sensation where the individual pulsations in speech can be distinguished (Titze, 1994). The threshold F0 for vocal creakiness was investigated by Keidar (1983) who presented listeners with synthetic vowel-like stimuli with different waveshapes and F0s between

40 Hz and 100 Hz. The waveshape of the stimuli was varied by manipulating, in the vowel synthesis, the open quotient (OQ) and the speed quotient (SQ) which equal the duration of the glottal open phase relative to the cycle length and the ratio between the glottal opening and closing phase, respectively. The likelihood of perceiving the stimulus as pulsating increased with decreasing F0, and at 70 Hz, the stimulus was judged as pulsating or continuous with equal probability. Thus, the lower limit of pitch perception seems to be higher for speech sounds with a pulse-like excitation than for stimuli such as RFN with a less pulse-like temporal envelope. Accordingly, Titze (1994) argued that the threshold for pulsation could depend, in addition to F0, on the decay time of the pulse so that a pulse-like perception could most easily be effected by stimuli where there are clear silent gaps between the excitation pulses. Contrary to this decay time hypothesis, Keidar (1983) reported that the threshold for pulse perception to be independent of the waveshape features OQ and SQ. However, even the vowel stimuli excited with the most gradual pulses that were used by Keidar (1983) probably had a more pulse-like temporal envelope than the RFN stimuli used by Guttman and Julesz (1963). Thus, a difference in the duration of the silent temporal gap within the fundamental period might explain why the threshold for pulse perception could be lower for RFN stimuli with no silent gaps (19 Hz; Guttman and Julesz, 1963) than for vowel-like stimuli with relatively rapid intraperiod decay envelopes (70 Hz; Keidar 1983). Nevertheless, the primary acoustic feature that determines whether the periodicity of a sound is perceived as a stable pitch or as a series of pulses is the F0.

The pitch of a sound can vary not only in the scale from low to high but also in its salience or strength (Fastl and Zwicker, 2007). While the scaling of pitch from low to high depends primarily on the sound F0, it could be argued that the primary acoustic feature contributing to pitch salience is the degree of sound periodicity. This view is intuitively apparent in that random noise, which lacks periodicity, elicits no pitch while the pitch of sounds such as speech and musical instruments with a high degree of periodicity can be easily perceived. The relationship between the degree of sound periodicity and pitch perception has been investigated in a more systematic way in several experimental studies (Yost, 1978, 1979; Yost *et al.*, 1978; Patterson *et al.*, 1996; Yost, 1996). The degree of sound periodicity in psychoacoustic experiments of pitch strength has often been manipulated by using iterated ripple noise stimuli. Ripple noise (RN) is generated by four steps which include 1) producing a random noise, 2) taking a copy of the noise, 3) delaying the copy of the noise in the time domain, and 4) adding this delayed copy to the original noise by means of linear summation. Delaying the copy of the noise before adding it to the original introduces quasiperiodic repetition in the

output signal where the period length (1/F0) is the delay time. When the delay-and-add procedure is repeated several times, iterated ripple noise (IRN) is produced. The degree of periodicity of an IRN stimulus increases as a function of the number of iterations used. Thus, the degree of periodicity (and the F0) of IRN can be manipulated in a systematic way. Perceptually, IRN resembles a combination of tonal pitch and noise, which are associated with the quasiperiodic and the random component of the stimulus, respectively (Patterson *et al.*, 1996). The strength of the pitch produced by IRN was investigated by Yost (1996) with a magnitude estimation procedure where the listeners assigned numerical values to indicate pitch salience. He showed that as the degree of stimulus periodicity (quantified as the height of the peak in the stimulus autocorrelation function that corresponds to the period length) increased, the judged pitch strength increased exponentially.

The pitch strength of an auditory test stimulus can be evaluated also through finding a reference stimulus that matches the pitch salience of the test stimulus. Patterson and colleagues (1996) studied the relative dominance of the tonal over the noisy component in the perception of IRN by using harmonic complex tones embedded in noise as reference stimuli. In particular, they used the psychophysical method of constant stimuli (*cf.*, Gescheider, 1997) where an IRN stimulus generated with a given number of iterations was compared against reference stimuli with variable tone-to-noise ratios. This procedure yielded psychometric functions that related the empirical probability of judging the reference stimulus as having a stronger pitch than the IRN stimulus to the tone-to-noise ratio of the reference stimulus. The point in the psychometric function at which the reference stimulus is judged as having a stronger pitch than the IRN stimulus on 50 % of the trials was considered as the point of subjective equality (PSE) where the pitch salience of the IRN and the reference stimuli were identical. This allowed the authors to operationalize the pitch strength of an IRN as the tone-to-noise ratio of the reference stimulus at the PSE. They reported an increase in the tone-to-noise ratio of the reference stimulus at the PSE, when the degree of IRN periodicity was increased. These results, thus, indicate a strong relationship between the degree of sound periodicity and pitch strength. However, the degree of periodicity is not the only factor that affects pitch salience, as the strength of the pitch produced by an auditory stimulus also depends on the F0, duration, and the bandwidth of the stimulus (Fastl and Zwicker, 2007).

Because sound periodicity is a feature which, by definition, is distributed over time, its detection requires temporal integration (Plack and Oxenham, 2005b). The temporal integration of periodicity underlying pitch perception can be described by a temporal window of integration (TWI) which determines the duration of a sound segment that affects the accuracy or the

salience of the elicited pitch. Human listeners are capable of perceiving the pitch of relatively short vowel stimuli (Robinson and Patterson, 1995) and are able to track changes in F0 (e.g, in intonational or tonal patterns of speech). The integration of periodicity must, therefore, be short enough to support the pitch perception of brief stimuli. Furthermore, the following of changes in F0 would conceivably require sampling of successive stimulus segments with a short integration window. The shortest duration of a stimulus (or its segment) that is required for the detection of periodicity can be termed as the minimal integration window of periodicity (Wiegrebe, 2001; Plack and Oxenham, 2005b). Then again, temporal integration facilitates the detection of periodicity over extended time windows as well. This aspect of temporal integration can be characterized with the maximal integration time, which is the longest stimulus duration up to which the detection of periodicity is enhanced (Plack and Oxenham, 2005b).

Temporal integration of periodicity was studied psychoacoustically by Wiegrebe (2001) with stimuli that contained alternating periodic and aperiodic segments. The assumption behind this approach is that the ability of a listener to detect an oscillation in pitch strength that corresponds to a modulation in the degree of stimulus periodicity can be used to estimate the TWI for periodicity. Accordingly, the TWI for periodicity can be identified as the shortest period length of alternation between periodic and aperiodic segments where the oscillation in pitch strength can be perceived. Wiegrebe (2001) measured the audibility of this oscillation for different rates of modulation in the degree of stimulus periodicity and for different F0s of the periodic segments. He further modeled the TWIs with exponentially decaying functions (x-axis: time, y-axis: weight) that weight stimulus periodicity over time. These functions were characterized by time constants that determine the duration of the stimulus segment over which periodicity is integrated. The perceptual data for stimuli with period lengths below 1.25 ms (800 Hz) could be explained with a uniform time constant of 2.5 ms. Importantly, for longer period lengths, the estimated time constant was twice the period length. Thus, the results of Wiegrebe suggest that the TWI for periodicity depends on the F0 of the stimulus so that the length of the TWI decreases with increasing F0 when the F0 is below 800 Hz. These estimates of periodicity integration, however, pertain only to the minimal integration time. Robinson and Patterson (1995) reported enhanced pitch recognition performance with increasing stimulus duration up to a duration of 32 cycles for vowel stimuli with F0s in the 33–345 Hz range. Thus, the maximal integration time for periodicity, related to pitch perception, can be up to one second in absolute time.

In summary, periodic sounds yield a perception of pitch with a specific height (from low to high) and salience (from weak to strong). The phenomena of pitch height, pitch salience, and temporal integration in pitch perception are interconnected and the perception of periodicity as pitch requires a favorable combination of the related sound features such as F0, duration, and the degree of periodicity.

3.2 Objective measures of sound periodicity

Formally, periodicity can be defined so that for a periodic signal $s(t)$, which has a period length $T > 0$, the equation $s(t) = s(t + T)$ is true for all values of t (de Cheveigné, 2005). Sound signals that do not conform to this definition are aperiodic. However, it is useful to recognize many natural sounds such as speech sounds as quasiperiodic when the sound approximately fits to the definition of periodicity (*e.g.*, Blauert and Xiang, 2009). Thus, the periodicity of a sound can be investigated in terms of estimating the period length T and the degree of sound periodicity, that is, the extent to which the sound is similar to the definition of a periodic signal. The F0 of a sound, which is the inverse of T, can be calculated objectively with time-domain and frequency-domain analyses from any periodic sound. Despite the apparent simplicity of finding the F0 of a periodic steady-state signal, the extraction of the F0 from speech and other quasiperiodic sounds may at times be difficult. Therefore, hundreds of different F0 extraction algorithms have been developed to cope with these difficulties (Hess, 2008). These algorithms can be categorized into time-domain algorithms and short-term analysis algorithms (Hess, 2008). In detecting the F0 of speech, the former are based on defining the period length as the elapsed time between successive laryngeal pulses or excitation cycles. Such time-domain extraction of speech periodicity can be facilitated with glottal inverse filtering or with other means of simplifications of the temporal structure of the signal. The latter algorithms estimate the F0 or its inverse from short-term frames derived from the sound signal. The short-term analysis algorithms can, further, be divided into correlation- and distance-based algorithms, cepstrum and other double-transform methods, frequency-domain harmonic analyses, active modeling using linear prediction, and into statistical methods.

Among the correlation-based approaches to F0 extraction, the use of autocorrelation analyses has been very popular (Rabiner, 1977) and is one of the oldest methods used in short-term analyses of periodicity (Hess, 2008). Autocorrelation is based on the product (covariance) between the signal and

its delayed version which lags the signal by a specific time-interval. Thus, autocorrelation can be expressed as a function of the lag. In such an autocorrelation function (ACF), a high value at a given lag indicates that the signal is similar to its delayed version and, thus, (quasi)periodic at the cycle length specified by the lag. The F0 can be estimated as the first maximum point in the ACF after the zero-lag delay (Fig. 3). Local maxima in the ACF of a periodic signal (beyond the lag corresponding to the fundamental period) can also be found at integer multiples of the fundamental period. The ACF of an aperiodic signal does not contain any prominent local maxima beyond the zero-lag point. The degree of sound periodicity can be calculated from the ACF as the height of the peak in the function at the lag corresponding to the fundamental period (Hess, 2008). In addition to the autocorrelation analyses, distance-based analysis (de Cheveigné, 1998) or the combination of distance-based and autocorrelation analyses (de Cheveigné and Kawahara, 2002) can used for the extraction of the F0 and the degree of periodicity by analyzing the signal as a function of the time-lag.

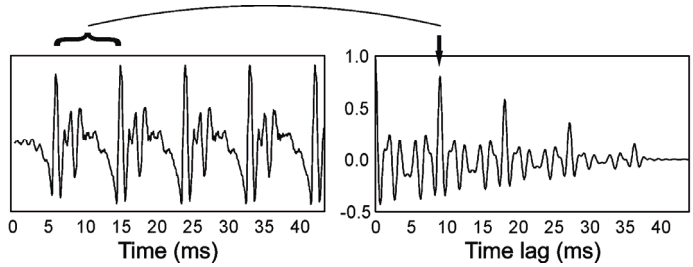

Figure 3. Detection of the fundamental period length (1/F0) with autocorrelation analysis. The waveform of a vowel sound [a] is shown on the left. The autocorrelation function (ACF), on the right, has the first maximum value (beyond the zero-lag point) at lag = 9 ms, which is the fundamental period length of the vowel (F0 = 111 Hz = 1 / 9 ms). A rectangular window was used in the calculation of the ACF.

The spectrum of a periodic sound has a regular comb structure where peaks can be found at integer multiples of the F0. Thus, the F0 can be extracted from the spectral comb structure of a sound. As the harmonic partials in the spectrum of a periodic sound are integer multiples of the F0, the F0 can be calculated as the greatest common divisor of the harmonics (Hess, 2008). Dividing the harmonic partials of a sound spectrum by integers produces so-called subharmonic spectra. The F0 can, then, be extracted by finding the maximum peak from the sum of the subharmonic spectra (Hermes, 1988; Fig. 4). The degree of sound periodicity, in turn, can be calculated from spectral

analyses with such metrics as the harmonics-to-noise ratio (*e.g.*, Lewis, *et al.*, 2009).

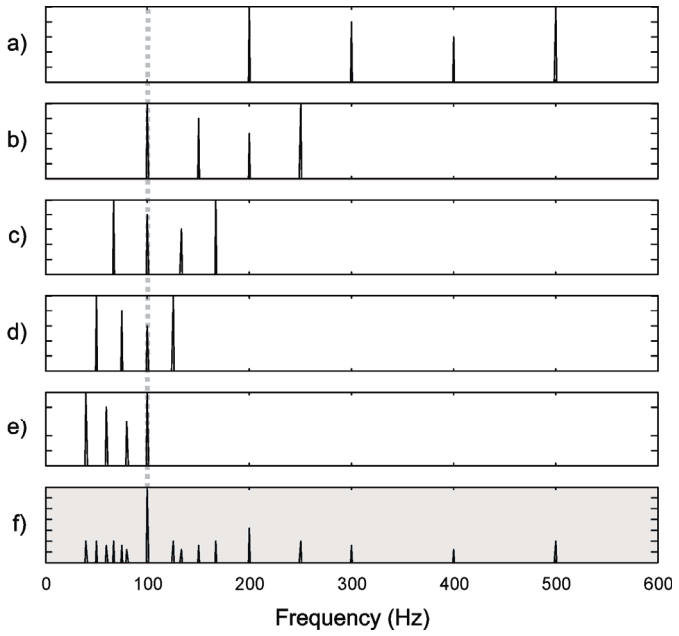

Figure 4. Detection of the F0 with subharmonic summation. The sound spectrum, shown in panel a), consists of four partials. The subharmonic spectra, shown in panels b–e), are produced by shifting the partials to lower frequencies by dividing the frequencies of the partials by integers 2–5, respectively. The spectra in panels a–e) are averaged into a summary spectrum which is shown in panel f). The most prominent partial in the summary spectrum [panel f)] is identified as the F0.

In conclusion, both autocorrelation and spectral subharmonic-based approaches of periodicity detection (among alternative approaches) can be used to estimate the F0 of a sound and to calculate the degree of sound periodicity. The algorithms for estimating the F0 of speech and other time-varying acoustic stimuli are used in application areas ranging from speech and music technology to basic linguistic research and clinical investigations of voice quality (Hess, 2008). Variations in the F0 and in the degree of periodicity of speech and other auditory stimuli produce significant changes in the perception produced by these stimuli (see Sections 2.1–2.3 and 3.1). In investigations of the perceptual aspects of sound periodicity, F0 extraction algorithms are important in characterizing the periodicity of the acoustic stimuli so that the relationships between the objective stimulus parameters of

sound periodicity can be related to subjective perceptions elicited by the stimuli. The research that aims to find the neural basis for the auditory processing of sound features is called auditory neuroscience. In auditory neuroscience, Fo extraction algorithms are important in both characterizing the periodicity of the acoustic stimuli and as elements of computational models of hearing. Thus, objective estimates of sound periodicity are necessary in understanding the neural representations of the Fo and the degree of sound periodicity.

4. Methods of auditory neuroscience in studies of speech perception

4.1 Overview of the methods used in auditory neuroscience

The ability of human listeners to encode acoustic features of speech in order to extract both the linguistic and the paralinguistic meaning of the speech utterances relies on the accurate representation of these features in the nervous system. Therefore, one of the most important tasks of auditory neuroscience is to uncover the way in which the features of speech are represented and processed along the auditory pathway from the cochlea to the cerebral cortex. The neuroscientific investigation of auditory perception comprises a wide array of methods including invasive electrophysiology, behavioral studies of both healthy listeners and of patients with brain damage (cognitive neuropsychology), as well as non-invasive hemodynamic, electric, and magnetic techniques.

The various methods for observing the activations in the brain have characteristic patterns of strengths and weaknesses (Baars and Ramsay, 2007). Invasive electrophysiology with single-unit recordings has a spatial resolution on the scale of individual neurons and a time resolution of around one millisecond which makes single-unit recordings the most precise of the techniques for studying brain activity. While the spatial resolution of the invasive methods surpasses that of non-invasive neuroimaging, neural activity in only a limited set of anatomical regions can be studied invasively at a time. Invasive single-unit measurements can, moreover, typically be obtained only from animal models which, in the auditory neuroscience of speech perception, may sometimes be problematic since the ability to encode the linguistic content of speech stimuli is unique to human subjects. While invasive multi-unit recordings from human patients undergoing surgery are sometimes possible with depth electrodes, the primary means of studying the activation elicited by speech sounds in the human brain is the use of non-invasive techniques.

The cortical processing of speech involves rapidly changing activity in distinct areas of the cortex. Therefore, its study requires both temporally and spatially accurate imaging techniques. The brain activity elicited by speech sounds can be non-invasively measured both with the hemodynamic methods of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) and with the electric and magnetic methods of electroencephalography (EEG) and magnetoencephalography (MEG).

In hemodynamic methods, the neural activity is indirectly inferred from signals that are related to aspects of regional blood flow (Griffiths, 2005) such as the blood oxygen level dependent (BOLD) signal. The major advantage of the hemodynamic imaging methods relative to the other techniques is the coverage of the whole brain which is combined with a high spatial resolution that can resolve volumes that are less than 1 cm in diameter. These methods, however, suffer from poor temporal resolution which is largely due to the slow variations in the hemodynamic response. Therefore PET and fMRI measurements cannot be used to follow rapid changes in neural activity. Hemodynamic methods are typically used to identify regions in the brain that are activated by a certain task or a stimulus feature. Such activated areas can be mapped by contrasting the hemodynamic signal derived from an experimental condition to that obtained from a control condition. This approach, however, leaves the interpretation of the observed active regions open to several alternatives (Huettel *et al.*, 2009). In particular, the observed activation may be due to an increased hemodynamic response in the experimental condition, a decreased response in the control condition, or a combination of these two possibilities. Furthermore, the site of the hemodynamic activation may differ from the site of the electric activity of the nerve cells.

The non-invasive electric and magnetic measurements with EEG and MEG, in turn, directly reflect the electric neural activation (Hämäläinen *et al.*, 1993). MEG and EEG are closely related in that both methods measure similar synchronized neural activity in the brain. However, while EEG measures the differences in electric potentials on the scalp produced by cerebral currents, MEG measures the magnetic fields produced by these currents. Both of these methods have excellent temporal resolution which is in the order of a millisecond. However, the localization of the activity that produces the EEG and MEG responses requires a number of assumptions. The electric signals originating from cerebral activity, measured with EEG, are distorted by the skull and the scalp. In contrast, the magnetic fields originating from cerebral cortex, measured with MEG, pass through the tissue layers undistorted. Therefore, fewer assumptions are needed in MEG source localization than in EEG source localization. In studies of auditory cortical activity, the lateral

location of the auditory regions may, further, lead to instability of source estimates from EEG data. Finally, in a comparison between the replicability of the dipolar source estimates from EEG and MEG data, the variability in EEG-based location coordinates was, at times, twice that of the MEG-based location coordinates (Virtanen *et al.*, 1998). MEG is used in many studies of the cortical processing of sound periodicity and in the studies of this thesis. Therefore, MEG will be described in more detail in the following chapter.

4.2 Magnetoencephalography (MEG)

MEG enables the investigation of neural activity by measuring the magnetic fields produced by the electric currents in the brain. These weak magnetic fields are recorded outside the head with ultrasensitive detectors called superconducting quantum interference devices (SQUIDs; Hämäläinen *et al.*, 1993). The SQUID arrays in contemporary MEG devices provide a coverage of the whole head so that activity in various parts of the brain can be measured simultaneously. The MEG system (Vectorview 4-D, Elekta Neuromag, Finland) employed in the studies of this thesis contains two types of sensors: planar gradiometers and magnetometers. The planar gradiometer signal patterns have a peak directly above underlying (focal) current sources and are, therefore, optimally suited for detecting cortical sources. Magnetometers have more widespread sensitivity patterns, especially in depth, and are thus more suitable for detecting fields arising from deeper structures (*e.g.*, Parkkonen *et al.*, 2009).

The principal sources of magnetic fields detected in MEG and also of the electric potentials recorded in EEG are the postsynaptic currents in the apical dendrites of cortical pyramidal cells rather than axonal currents. This is partly due to the different durations of the postsynaptic potentials and the action potentials (Lopes da Silva, 2010) which are in the order of 10 ms or more and 1 ms, respectively (Hämäläinen *et al.*, 1993). Thus, the probability of temporally overlapping activity is higher for the slow postsynaptic potentials than for the rapid action potentials (Lopes da Silva, 2010). Moreover, the magnetic fields produced by the axonal currents attenuate more strongly with distance than the magnetic fields produced by the postsynaptic currents.

For the magnetic signal to be detectable with MEG, synchronous activity in a large number of neurons is required (Hämäläinen *et al.*, 1993). It is estimated that the weakest cortical signals which can be measured with MEG, are in the order of 10 nAm (Hämäläinen *et al.*, 1993). Accordingly, if the strength of a

current dipole of a cortical pyramidal cell is around 0.2 pAm (Murakami and Okada, 2006) approximately 50000 neurons must be simultaneously active to produce a magnetic field measurable with MEG (Lopes da Silva, 2010). The neurons need also to be oriented in parallel so that the net macroscopic current and, thus, a detectable magnetic field is obtained. The orientation of the neural currents is also crucial as the currents tangential to the head produce magnetic fields outside the head while the radial currents cannot be detected. Finally, the strength of the MEG signal decreases as a function of the distance of the active area of the brain from the sensors. The neural populations that best conform to the above requirements in the auditory areas of the temporal lobes consist of the pyramidal cells located in cortical fissures. In practice, the minimum cerebral activity that can be detected with MEG depends on the acquisition paradigm, on the number of averaged epochs, the MEG analysis methods, and the statistical analyses performed on the data in addition to the properties of the neural generators and the MEG sensors.

The spatial resolution of MEG is limited by the pooling of the neural activity from extended areas to the recorded magnetic field patterns. Furthermore, localization of the active regions is an inverse problem with, in principle, an infinite number of solutions (Baillet, 2010). The localization of the brain areas that are most probably active is, nevertheless, possible with analyzing the MEG data with computational models (Hämäläinen *et al.*, 1993). A widely used method for localizing the sources of MEG responses is the equivalent current dipole (ECD) model. In this approach, the magnetic field pattern detected by the MEG sensors is explained with one or more current dipoles with specific locations and orientations. This focal source model can adequately explain the magnetic field pattern produced by transient event-related auditory cortical activation. Since the auditory activity is typically bilateral, two current dipoles, one in each hemisphere, are typically needed to explain the MEG data. However, due to the large distance between these sources the two dipoles can be usually fitted independently of each other. Alternative approaches to the source modeling of MEG responses include the estimation of distributed sources with minimum-norm estimates (Hämäläinen *et al.*, 1993; Hämäläinen and Ilmoniemi, 1994) or minimum-current estimates (Hämäläinen *et al.*, 1993; Uutela *et al.*, 1999).

In conclusion, MEG provides means to measure neural activity elicited simultaneously in different parts of the cerebral cortex. Of particular relevance to the investigations of speech perception is the sensitivity of MEG to the magnetic fields produced in the auditory cortical areas. Due to the high temporal resolution of the method, it is also possible to follow rapid changes in neural activity underlying these MEG responses. The spatial resolution of the method, moreover, allows localizing the underlying neural generators of

the magnetic fields and enables, for example, investigating the activity in the left- and the right-hemispheric cortical areas separately. Finally, MEG is totally non-invasive and thus permits the measurements of neural activity elicited by speech stimuli in the brains of healthy human subjects.

4.3 Auditory evoked fields (AEFs)

The cortical activity that can be measured with MEG and EEG contains both spontaneous ongoing activity and transient events elicited by sensory stimuli (Shah *et al.*, 2004; Mazaheri and Jensen, 2006). The single transient responses of the brain elicited by auditory stimuli are of about the same amplitude as the simultaneous ongoing activity in the nearby cortical areas and are, thus, difficult to measure with MEG. However, the transient responses are time-locked to the presentation of an auditory stimulus and their signal-to-noise ratio can be increased by averaging the signals over multiple repetitions of the stimuli. In averaging, the brain activity that is not related to and, hence, not time-locked to the auditory stimulus will be canceled out while the activity related to the (recurring) stimulus event will be emphasized. Thus, reliable measurements of auditory cortical activity can be obtained from averaged MEG data.

The magnetic fields recorded as the averaged responses elicited by repeated auditory stimuli are called auditory evoked fields (AEFs; Figure 5). A short auditory stimulus elicits a well-defined P1m-N1m-P2m wave which is a magnetic counterpart of the P1-N1-P2 wave recorded with EEG. This wave consists of a series of responses elicited around 50 ms, 100 ms, and 200 ms after stimulus onset (May and Tiitinen, 2010). While the N1(m) can be elicited by any audible stimulus, the sustained field (SF) or sustained potential (SP) in EEG, which reaches its maximum around 400 ms, is elicited only by stimuli with a prolonged duration (Picton *et al.*, 1978a, 1978b; Hari *et al.*, 1980; Pantev *et al.*, 1994). The N1(m) is generated by multiple sources located in Heschl's gyrus (HG) and in the superior temporal gyrus (STG) (for a review, see May and Tiitinen, 2010). The generators of the N1(m) response are highly sensitive to many acoustic features of auditory stimuli. This is reflected by stimulus-dependent modulations in the amplitude, latency, and source location of the response. In particular, the acoustic features that are crucial for speech perception such as vowel identity and the signal-to-noise ratio of speech sounds are reflected in the above characteristics of the N1(m) response (Diesch *et al.*, 1996; Martin *et al.*, 1997, 2005; Obleser *et al.*, 2003a, 2003b, 2004; Roberts *et al.*, 2004; Mäkelä *et al.*, 2004, 2005; Tiitinen *et al.*, 2005;

Liikkanen *et al.*, 2007; Miettinen *et al.*, 2011). Thus, the auditory cortical activity related to the processing of speech sounds can be investigated through the N1m response.

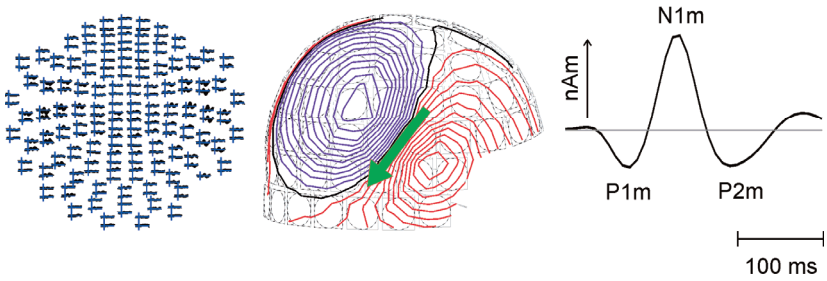

Figure 5. Measurement of an auditory evoked field (AEF) with magnetoencephalography (MEG). The magnetic field produced by the electric activity in the brain can be measured with an array of 204 gradiometer sensors using a whole-head MEG device (Vectorview 4-D, Elekta Neuromag, Finland) as shown on the left. A typical pattern of the magnetic field elicited around 100 ms after the onset of an auditory stimulus is shown in the middle. Areas surrounded with red and blue contour lines show the outflux and influx of magnetic fields, respectively. The observed auditory N1m response can be explained with an equivalent current dipole (ECD, depicted with an arrow) located in the temporal lobe areas of the cerebral cortex. The variation in the strength of the ECD over time can be investigated from a source waveform as shown on the right.

Although a difference in the source location of the N1m and the SF has been observed, both responses are generated by feature selective neural populations in the supratemporal cortex (Pantev *et al.*, 1994). Further, the feature-selectivity of the underlying neural generators of the SF in differentiating between speech and non-speech stimuli has been implicated by the results of Hewson-Stoate and colleagues (2006) as well as by the results of Gutschalk and Uppenkamp (2011). Interestingly, a recent study by Gutschalk and colleagues (2010) suggested that the N1m, but not the SF response, was tightly coupled to the BOLD signal measured with fMRI. Therefore, it seems that MEG has an advantage over fMRI in capturing the sustained aspects of cortical auditory-evoked activity. Due to the high temporal resolution of MEG, the N1m and the SF responses can be studied independently with suitable stimulation paradigms. Given that the generators of the SF are sensitive to the cortical processing of speech sounds, investigations of such sustained responses with MEG may provide important additional data to complement the hemodynamic research on speech perception.

Although the amplitude, latency and source location of AEFs allow comparisons of cortical activity elicited by distinct stimuli, their

interpretations are not straightforward. Firstly, the AEF amplitude can depend on several variables such as the number of active neurons, the synchronization within the neural population activated, the location of the MEG sensor array with respect to the active region, and the cancellation effects between the simultaneously active generators (Ahlfors *et al.*, 2010; Lopes da Silva, 2010). Secondly, while a best fitting source estimate to explain the observed AEF can be found, the precise characteristics of this source estimate will always depend *a priori* on the localization approach (Hämäläinen *et al.*, 1993; Mäkelä, 2006). Thus, measurements of the cortical activity with AEFs are often open to various interpretations. Therefore, relating the findings from investigations of AEFs to data from hemodynamic and animal models is important in constraining the interpretations that can be made from such findings.

4.4 Stimulus-specific modulations of the AEFs

Although AEFs reflect the activity of large neural populations (Hämäläinen *et al.*, 1993), the role of specific populations to the elicitation of an AEF can be extracted for more detailed analyses with suitable experimental designs. That is, the contribution of distinct neural populations can be inferred from the modulations in the AEFs across different experimental conditions. An experimental condition in an event-related MEG study is often defined as a stimulus sequence where a single isolated stimulus is repeated several times (*e.g.*, more than 100 times). Thus, changes in the amplitude, latency, or source location of the AEFs elicited by variable stimuli in separate sequences can be taken to suggest that the pattern of neural activity differed between the stimulation conditions.

The cortical feature-sensitivity can also, under favorable conditions, be revealed through the paradigm of stimulus-specific adaptation (SSA; Butler, 1968; Jääskeläinen *et al.*, 2004; Ahveninen *et al.*, 2006; Salminen *et al.*, 2009, 2010; May and Tiitinen, 2010). The paradigm is based on the adaptation, that is, the attenuation of cortical activity with repeated presentation of a stimulus. Cortical adaptation is stimulus-specific in the sense that the attenuation (of *e.g.*, the N1m response) is largest when the acoustic features of the stimuli are held uniform throughout the stimulus sequence. Conversely, this attenuation is reduced when the stimuli presented in the sequence are acoustically variable. The stimulus-specificity of adaptation is thought to arise from the feature-selectivity of different cortical populations and from synaptic depression so that the activity of a specific population is attenuated only if the

sequence of stimuli matches the preferred input of the population (*cf.*, Ulanovsky *et al.*, 2003, 2004). Cortical adaptation, furthermore, decays as a function of time which is reflected by the increase in amplitude of the N1m response as a function of the interstimulus interval (ISI) (Lu *et al.*, 1992).

In the SSA paradigm, the effects of an adaptor stimulus on the response elicited by a subsequent probe stimulus are investigated. In practice, adaptors and probes, in a given experimental condition, are alternated repeatedly in a stimulus sequence. The attenuation of neural activity caused by a specific adaptor is measured with reference to the condition where the adaptor is identical to the probe (and yields an activity pattern that best matches that elicited by the probe). Importantly, if there is a release from adaptation for a specific adaptor-probe combination against the reference condition, the probe can be suggested to activate, at least partly, a different neural population than the adaptor.

In conclusion, stimulus-specific tuning of the auditory cortex can be potentially revealed by changes in AEFs that are related to experimental manipulations of stimulus features. The effects of stimulus features on AEFs can be investigated either by directly comparing the AEFs elicited in separate stimulus sequences or by using the SSA paradigm. The interpretation of stimulus-specific modulations in the AEFs obtained by direct comparison of AEFs elicited in separate sequences requires fever assumptions. Furthermore, the stimulus-specificity of the AEFs, elicited in experimental conditions where a single isolated stimulus is repeated, has been observed in a vast range of empirical studies (for a review, see May and Tiitinen, 2010). However, in certain cases, the SSA paradigm may be more sensitive to cortical tuning to stimulus parameters (Salminen *et al.*, 2009, 2010). Independent of the approach employed, modulations in the AEFs produced across experimental conditions are suggestive of cortical selectivity to the experimentally varied feature of the auditory stimuli.

5. Neural representations of sound periodicity

5.1 Periodicity-sensitive activity in the auditory cortex

In the auditory pathway, the cortical level is preceded by several anatomical structures where the periodicity of sound is represented in different ways. The decrease in the upper limit of phase locking along the ascending pathway seems to require a conversion of the representation of sound periodicity from a temporally accurate phase-locked copy of the stimulus to a topographic or otherwise feature-selective representation (Lu *et al.*, 2001; Winter, 2005). In an fMRI study, Griffiths and colleagues (2001) identified cochlear nucleus (CN) as the first anatomical site that yields an activation contrast between periodic and aperiodic stimulation conditions. This periodicity-specificity of the hemodynamic response in the CN, however, has been thought to reflect neural synchronization induced by the stimulus periodicity rather than the activation of periodicity-specific neurons as such. That is, the same neurons of the CN are probably activated by both periodic and aperiodic stimuli (Griffiths *et al.*, 2001; Griffiths, 2005).

In the inferior colliculi (IC), the contrast between the periodic and aperiodic conditions is markedly stronger than in the CN and can hardly be explained by neural synchronization only. Thus, hemodynamic imaging provides evidence for the encoding of sound periodicity with feature-selective populations in IC (Griffiths *et al.*, 2001). These results are in line with the anatomically oriented model of periodicity processing by Langner (1992) that proposes that the IC is the first stage in the auditory pathway to contain periodicity-sensitive neurons. The results indicating periodicity-sensitivity in subcortical structures of CN and IC were, however, not replicated in the study by Penagos and colleagues (2004) who argue that the previous findings by Griffiths and colleagues (2001) may have been due to a confounding influence of auditory distortion products.

The periodicity-sensitivity observed at the level of the IC is preserved at the cortical level as indicated by several fMRI and PET studies (Griffiths *et al.*,

1998, 2001; Patterson *et al.*, 2002; Penagos *et al.*, 2004, Hall *et al.*, 2005, 2006; Hall and Plack, 2009; Lewis *et al.*, 2009; Garcia *et al.*, 2010; von Kriegstein *et al.*, 2010). Periodicity-specific activation has been localized to anterior areas of the nonprimary auditory cortex (Griffiths *et al.*, 2001) more specifically identified as the lateral Heschl's Gyrus (lHG) (Patterson *et al.*, 2002) with some spread onto the STG in the left hemisphere (Penagos *et al.*, 2004).

The most commonly used pitch-evoking stimulus in hemodynamic studies of cortical processing of sound periodicity has been the IRN (*cf.*, Griffiths *et al.*, 1998, 2001; Patterson *et al.*, 2002; Hall *et al.*, 2005, 2006; Hall and Plack 2009; Lewis *et al.*, 2009). Interestingly, the studies using other stimulus types have demonstrated activation patterns that are different from those elicited by IRN stimuli (Hall and Plack, 2009; Lewis *et al.*, 2009; Garcia *et al.*, 2010). In particular, the view that lHG constitutes the center of periodicity processing in the cortex (*e.g.*, Patterson *et al.*, 2002) was recently disputed by Hall and Plack (2009) who used a broader range of periodic stimuli, including harmonic complex tones with variable bandwidths, in an fMRI study. They suggested that parts of the planum temporale (PT) are more relevant for periodicity processing than lHG and reported further periodicity-sensitive activation in areas such as the temporo-parieto-occipital junction and the prefrontal cortex of some of their subjects. Concurring with the findings of Hall and Plack (2009), Garcia and colleagues (2010) found the most significant pitch-related activity to be centered in the posterior auditory cortex, in lateral PT. In addition, using animal vocalizations as stimuli, Lewis and colleagues (2009) observed periodicity-sensitive activation along the medial STG that was largely non-overlapping with the periodicity-specific activation elicited by the IRN stimuli. Thus, rather than being focused in a well-defined center, the periodicity-specific regions of the cortex seem to be distributed over several anatomical areas. Moreover, the extent and the location of the periodicity-specific activation seem to depend on the type of periodic stimuli, rather than being similar for all periodic sounds. Finally, Hall and Plack (2009) suggested that features of IRN that are not related to periodicity might affect the localization results. Thus, the location and extent of periodicity-specific activation in the auditory cortex are under dispute.

The cortical processing of sound periodicity has been investigated in several MEG studies as well. The results from these MEG studies have provided further evidence for the sensitivity of the auditory cortex to sound periodicity. The sensitivity to periodicity in many of the MEG studies has been operationalized as the differences in the AEFs elicited by periodic and aperiodic stimuli. The results from such studies indicate that the sensitivity to sound periodicity is manifested in both the amplitude and the source location

of the N1m response. In particular, the amplitude of the N1m elicited by periodic sounds is larger than that of the N1m elicited by aperiodic sounds (Hertrich *et al.*, 2000; Alku *et al.*, 2001; Soeta *et al.*, 2005; Tiitinen *et al.*, 2005; Lütkenhöner *et al.*, 2006). The source of the N1m in the case of periodic stimuli is, further, anterior to the source of the N1m elicited by aperiodic stimuli (Alku *et al.*, 2001; Tiitinen *et al.*, 2005). The increased amplitude of the N1m response in the periodic condition might be due to an increased number of neurons that are activated by periodic in contrast to aperiodic stimuli. The shift in the source location of the N1m between the periodic and the aperiodic condition, in turn, suggests that the spatial distribution might be different between the patterns of activity elicited by periodic and aperiodic sounds. Importantly, these results might be interpreted in the light of the hemodynamic findings of cortical processing of sound periodicity (Griffiths *et al.*, 1998, 2001; Patterson *et al.*, 2002; Penagos *et al.*, 2004, Hall *et al.*, 2005, 2006; Hall and Plack, 2009; Lewis *et al.*, 2009; Garcia *et al.*, 2010; von Kriegstein *et al.*, 2010) to suggest that periodic sounds activate a distinct periodicity-specific neural population in the auditory cortex.

As discussed above (please, see Section 4.4), an alternative approach in MEG and EEG is to use the SSA paradigm. In studies of cortical processing of sound periodicity, a version of the SSA paradigm has been used where a continuous or prolonged sound changes from aperiodic to periodic (or vice versa) without an intervening silent gap. One of the first of such studies was conducted by Kaukoranta and colleagues (1987) who presented subjects with syllable sounds that changed from an unvoiced fricative to a voiced vowel sound. In addition to the N1m elicited by the syllable onset, the authors observed a second N1m elicited by the onset of sound periodicity after noise, that is, at the consonant-vowel transition. The elicitation of an N1m, or its EEG-equivalent N1, by the onset of periodicity after an aperiodic stimulus segment has been subsequently observed by several authors (Mäkelä *et al.*, 1988; Martin and Boothroyd, 1999; Krumbholz *et al.*, 2003; Gutschalk *et al.*, 2004; Lütkenhöner *et al.*, 2011). An interpretation of these results based on the SSA approach suggests that the N1(m) elicited by the change from aperiodic to periodic stimulation reflects the activation of new periodicity-sensitive neural units not adapted by the aperiodic stimulation. Furthermore, the source location of the N1m response elicited by the onset of sound periodicity was more anterior than the source location of the N1m elicited by the onset of an aperiodic stimulus after silence (Kaukoranta *et al.*, 1987; Gutschalk *et al.*, 2004). This difference in the source location between the stimulation conditions, moreover, suggests a distinction between the neural generators giving rise to the N1m response in the case of periodic as opposed to aperiodic stimulation. Among the studies using the SSA paradigm (Mäkelä *et al.*, 1988;

Martin and Boothroyd, 1999; Krumbholz *et al.*, 2003; Gutschalk *et al.*, 2004; Lütkenhöner *et al.*, 2011), an N1 elicited by the onset of aperiodic noise after a periodic sound was observed only by Martin and Boothroyd (1999). Interestingly, the results of Martin and Boothroyd (1999) suggest the activation of an aperiodicity-sensitive population elicited by the noise stimulus which is a result consistent with recent fMRI findings of von Kriegstein and colleagues (2010) indicating an activation contrast between whispered and voiced syllables in the posteromedial HG. In conclusion, the SSA approach strongly supports the view that sound periodicity is processed in the auditory cortex by a dedicated neural population. However, the SSA data provide inconsistent evidence for the existence of a corresponding aperiodicity-sensitive neural population.

The localization of the periodicity-specific activity in the human brain, without the need to make extensive assumptions as in non-invasive hemodynamic and electromagnetic studies, is possible with depth electrode recordings. Such data corroborate the cortical sensitivity to the degree of sound periodicity revealed previously in the non-invasive studies (Schönwiesner and Zatorre, 2008; Griffiths *et al.*, 2010). According to Schönwiesner and Zatorre (2008), the periodicity-sensitive region resides in the lateral end of the superior temporal plane around HG as revealed by the activity elicited by the onset of periodicity after noise. Griffiths and colleagues (2010), however, found periodicity-sensitive activity in terms of induced high gamma-band oscillations in predominantly medial parts of the HG. Thus, human intracerebral recordings indicate periodicity-sensitive activity in terms of both transient and oscillatory responses in the auditory cortical areas. Given the limited subject count of one (Schönwiesner and Zatorre, 2008) and two (Griffiths *et al.*, 2010) patients, the differences in the location of the periodicity-specific region might be partly accounted for by the anatomical variability in the HG between the subjects (Griffiths *et al.*, 2010). Nevertheless, together with the hemodynamic results, these data might suggest that the processing of sound periodicity is spatially distributed in the auditory cortex.

A definitive mapping of the cortical representations of periodicity on the single cell level has not been very successful (Winter, 2005). Recent electrophysiological findings in marmosets by Bendor and Wang (2005, 2010), however, provide evidence for periodicity-selective neurons in the auditory cortex. The periodicity-specific activity on the single cell level was initially revealed by the tuning of a limited number of neurons to the F0 of an auditory stimulus (Bendor and Wang, 2005, 2010). This tuning to F0 was observed irrespective of whether the spectral harmonic at the F0 was present or not, that is, for the periodicity of the stimuli with a missing fundamental

frequency as well. The authors confirmed that part of such neurons were also sensitive to the degree of sound periodicity (Bendor and Wang, 2005, 2010). Measuring the single- and multi-unit activity elicited by synthetic vowels of different F0s in several cortical fields of ferrets, Bizley and colleagues (2010) were unable to find reliable level-independent representations of the stimulus F0 in individual neural discharge patterns. Analyzing the data with a simple classifier, however, revealed that the ensembles of neural activity patterns in the studied cortical fields supported an accuracy in the discrimination of stimulus F0 that matched the animals' behavioral performance. Thus, in the auditory cortex, sound periodicity might be to a considerable extent be encoded by a distributed representation rather than by the tuning properties of individual neurons.

In summary, evidence from both invasive and non-invasive studies of the human brain suggests the existence of periodicity-sensitive population in the auditory cortex. In particular, it seems that this population is located anterolateral to the primary auditory areas and that the net activity in the auditory cortex may be stronger in the case of periodic than in the case of aperiodic stimulation. However, the representations of sound periodicity in the single cell level and in the case of aperiodicity-sensitive activity seem to be unresolved.

5.2 Hemispheric lateralization of the representations of sound periodicity

It is widely accepted that the processing of spoken language is lateralized to the left hemisphere of the brain. Accordingly, the extraction of linguistic features from speech sounds would take place predominantly in the left-hemispheric regions of the brain as is evidenced by human neuroimaging (PET and fMRI) studies (*e.g.*, Phillips and Farmer, 1990; Belin *et al.*, 1998; Jäncke *et al.*, 2002; Zaehle *et al.*, 2004). The view of left-lateralized processing of speech is often complemented with the conception of right-lateralized processing of sound periodicity (Zatorre, 2001). Cognitive neuropsychological data corroborate the right-lateralization of the cortical populations that are responsible for the pitch perception of periodic sounds. Zatorre (1988) studied the ability of patients with unilateral temporal-lobe excisions to discriminate between harmonic stimuli where the partial corresponding to the F0 was either present or missing from the sound spectrum. All patients were able to discriminate well between the stimulus frequencies when the F0 was present in the sound spectrum. The patients with

right-hemispheric damage in the HG, however, made significantly more errors than healthy listeners in discriminating the frequencies of the stimuli with missing fundamentals. Comparable left-hemispheric excisions did not affect the discrimination between the missing fundamentals.

However, Tramo (2005), pointed out that the frequency differences between the stimuli used in the discrimination task by Zatorre (1988) were overly large in comparison to the normal thresholds of frequency discrimination. The frequency difference between the stimuli that was to be discriminated in the study by Zatorre (1988) was 40 % relative to the mean frequency of the stimulus pair. This frequency difference is considerably larger than the normal frequency discrimination threshold which is around 1 % relative to the mean frequency of a stimulus pair (Tramo, 2005). Thus, the frequency difference used by Zatorre (1988) was probably too large to permit a sensitive evaluation of the impact of left-hemispheric damage on the ability to discriminate between sound frequencies. However, the differential effect of right- as opposed to left-hemispheric damage in HG for missing fundamental discrimination suggests a lateralization of certain aspects of the representation of sound periodicity to the right hemisphere. Furthermore, damage to the right but not to the left HG has been subsequently shown to decrease, but not completely to remove, the ability to detect the direction of change in F0 (Johnsrude *et al.*, 2000). The simple discrimination of sound periodicity, where the perception of the direction of frequency change was not required, remained unimpaired in patients with right HG damage studied by Johnsrude and colleagues (2000). A critical role for the right-hemispheric auditory areas for frequency discrimination (Sidtis and Volpe, 1988; Divenyi and Robinson, 1989; Robin *et al.*, 1990) and frequency matching (Robin *et al.*, 1990) has been observed in several other cognitive neuropsychological studies of brain damage as well.

The lateralization of the processing of sound periodicity to the right hemisphere has been further tested with neuroimaging studies from healthy subjects. These studies have indicated right-lateralized activation in secondary auditory areas for conditions of passive listening to melodic patterns of periodic sounds (Patterson *et al.*, 2002), for increased pitch-related echoic memory load (Zatorre *et al.*, 1994; Zatorre, 2001), for the maintenance of pitch while singing (Perry *et al.*, 1999), and for imaging tunes (Halpern and Zatorre, 1999). Additionally, the activation in the right- but not in the left-hemispheric auditory areas has been shown to increase as a function of F0 distances within melodic sequences (Hyde *et al.*, 2008).

While many important aspects of the processing of sound periodicity may be lateralized to the right-hemispheric auditory areas, the left-hemispheric areas are likely to participate in the processing of sound periodicity as well. Firstly,

hemodynamic studies have consistently indicated bilateral activation contrasts between periodic and aperiodic conditions when the F0 of the periodic stimuli has been kept constant (*e.g.*, Griffiths *et al.*, 1998, 2001; Patterson *et al.*, 2002). While a condition with variable F0s produced stronger activation in the right than in the left auditory areas in the study by Patterson and colleagues (2002), left-hemispheric activation was observed by these authors and by Griffiths and colleagues (1998, 2001) who used comparable experimental conditions. Secondly, the periodicity-specific enhancement in the amplitude of the AEFs (Hertrich *et al.*, 2000; Alku *et al.*, 2001; Gutschalk *et al.*, 2004; Soeta *et al.*, 2005; Tiitinen *et al.*, 2005) and the anterior shift in the source location of the AEFs in conditions with periodic as opposed to aperiodic stimuli (Alku *et al.*, 2001; Gutschalk *et al.*, 2004; Tiitinen *et al.*, 2005) have been consistently observed bilaterally and symmetrically in the left and the right cortical hemisphere. Thus, there is convincing evidence of the contribution of the left-hemispheric auditory areas to the processing of sound periodicity. In particular, the representation of the periodicity of individual sounds, in contrast to (pseudo)melodic patterns, appears to be distributed symmetrically across the cortical hemispheres.

Schneider and colleagues (2005) showed in their MEG study, that the hemispheric lateralization of pitch perception, indexed by the amplitude of the P1m response depended on the perceptual strategy used by the subjects. Interestingly, a larger P1m was observed in the left than in the right hemisphere when the subjects perceived the pitch based on the missing fundamental rather than on the lowest spectral harmonic. The anatomical magnetic resonance images of the subjects with a tendency to perceive the pitch of the missing fundamental also revealed an increased volume of the gray matter in the left lHG. The results of Schneider and colleagues (2005), thus, seem to be at odds with the cognitive neuropsychological findings of Zatorre (1988) which suggest a right-lateralized processing of the pitch of the missing fundamental.

In conclusion, there is evidence for the lateralization of function of some aspects of the processing of sound periodicity into the right hemisphere of the brain. Such lateralization seems to characterize especially the processing of differences in F0 between two or more successive sounds. A bilaterally symmetric representation of the periodicity of individual sounds is, however, suggested by MEG and hemodynamic studies.

5.3 Relationships between the auditory cortical processing of speech and periodicity

To date, hemodynamic studies on cortical processing of the periodicity of speech sounds have been rare in comparison to the studies using non-speech stimuli in the investigations of cortical periodicity-specific activation. As an exception, von Kriegstein and colleagues (2010) investigated the cortical activation elicited by both periodic voiced and aperiodic whispered syllables. The cortical area producing an activation contrast between the voiced and the whispered condition was, according to the authors, located in the anterolateral HG. This site has been observed to produce a hemodynamic contrast between the periodic and the aperiodic stimulus conditions in several earlier studies using non-speech sounds (Griffiths *et al.*, 2001; Patterson *et al.*, 2002; Penagos *et al.*, 2004). The activation map for the voiced against whispered contrast in the group mean results of von Kriegstein and colleagues (2010), however, appears to have a significant spread into PT and planum polare (PP) as well. In this respect the results of von Kriegstein and colleagues were perhaps more akin to those of Garcia and colleagues (2010) who suggested that PT might be especially important in the cortical processing of sound periodicity. Von Kriegstein and colleagues (2010) observed a significant activation contrast in the reverse direction also, that is, between the aperiodic condition (whispered syllables) and the periodic condition (voiced syllables). The cortical area producing this aperiodicity-specific activation contrast was located in the posteromedial HG. A comparable pattern of activation in terms of an aperiodic-against-periodic contrast has not been found in the hemodynamic studies using non-speech studies (*e.g.*, Griffiths *et al.*, 2001; Patterson *et al.*, 2002; von Kriegstein *et al.*, 2006). Importantly, the aperiodicity-selective activation contrast elicited by the speech stimuli as observed by von Kriegstein and colleagues (2010) might suggest the existence of an aperiodicity-sensitive neural population in the auditory cortical areas.

A combined investigation of cortical activity related to speech and to sound periodicity with MEG was made by Gutschalk and Uppenkamp (2011). They produced stimuli that could be independently labeled as either “vowel” or “non-vowel” based on their resemblance to speech (speech vs. non-speech) and as either “periodic” or “non-periodic” based on the degree of stimulus periodicity (periodic vs. jittered). The speech-like stimuli were synthesized vowel sounds characterized by a steady-state formant structure whereas the non-speech stimuli contained rapid cycle-to-cycle variations in the formant structure. Consequently, the synthesized speech sounds could be identified as vowels whereas the non-speech sounds could not. The degree of stimulus periodicity, in turn, was manipulated by shifting the starting point of the

speech pulses in a cycle-to-cycle basis by a random amount of time between -6 ms and 6 ms. Thus, four classes of stimuli emerged from the manipulations: (1) periodic vowels (2) non-periodic vowels, (3) periodic non-vowels, and (4) non-periodic non-vowels. The cortical activity related to the periodicity and the speech-likeness of the stimuli was investigated from AEFs elicited in conditions where either periodicity or an identifiable formant structure was present, respectively, and by subtracting the AEFs elicited in conditions where these qualities were absent (or corrupted) from the former AEFs. This method combined with ECD modeling yielded bilaterally three distinct source estimates for the SF response termed "pitch", "vowel", and "unspecific" by the authors. These sources were obtained by fitting ECDs to MEG waveforms derived from contrasts between all periodic and all jittered conditions, all vowel against non-vowel conditions, and from the jittered non-vowel condition, respectively. Both the "pitch" source and the "vowel" source were located in the lateral HG. Based on these findings, Gutschalk and Uppenkamp (2011) suggested that the periodicity of both speech and non-speech sounds is processed in the same cortical area which also participates in early vowel-specific processing. Interestingly, they also found an effect of periodicity in the "vowel" source and an effect of speech-likeness in the "pitch" source. In particular, the activity in the "vowel" source was larger for the periodic than for the jittered vowels and the activity in the "pitch" source was larger for the vowel-like than for the non-speech stimuli. These results, thus, suggest that the cortical periodicity-specific activity might be different between speech and non-speech sounds, despite the site of such periodicity-specific activity being close to that of vowel-specific activity.

One of the major goals of invasive studies of auditory processing of periodicity in animal models has been the identification of neural units that provide a representation of sound periodicity independent of other sound features such as timbre, location, and loudness. Bizley and colleagues (2009) investigated the activity of cortical neurons of ferrets elicited by synthetic vowel stimuli that were varied in F0, timbre, and location. Most (65 %) neurons seemed to be modulated by two or more stimulus features and the neurons that were sensitive to only one stimulus feature were modulated by sound timbre. While these results alone do not preclude the existence of neurons that are modulated only by sound periodicity, they suggest that neural activity in many periodicity-sensitive units may also reflect other stimulus features besides periodicity.

In conclusion, the cortical activity elicited by sound periodicity may be dependent on other features of auditory stimuli besides periodicity. Furthermore, different results of cortical periodicity-specific activity may be obtained when using speech as opposed to non-speech stimuli. Thus, the

cortical processing of the periodicity of speech sounds may be revealed accurately only by using representative speech sounds stimuli.

5.4 Representation of the fundamental frequency of periodic sounds in the auditory cortex

The pitch of a sound, on a scale from low to high, is determined by the F0 of the sound. Therefore, a comprehensive understanding of the cortical processing of sound periodicity must cover the representations of not only the presence or absence of periodicity but the way in which the F0 of periodic sounds is represented in the cortex. The MEG results of Pantev and colleagues (1989) and Langner and colleagues (1997) suggest that the F0 of periodic sounds is mapped in the cortex with topographic organization where sounds with adjacent F0 values activate adjacent neural populations. Langner and colleagues (1997) observed orthogonal tonotopic and periodotopic representations of sound frequency and F0, respectively, in auditory cortex. The results of Pantev and colleagues (1989), on the contrary, suggest that the tonotopic and periodotopic representations of F0 run in parallel. These findings of periodotopy were, however, not replicated by Lütkenhöner (2003), who argued that the ECD method for source localization used in the above studies is unsuitable for the investigations of cortical tonotopy or periodotopy (Lütkenhöner, 2003; Lütkenhöner *et al.*, 2003). In particular, despite excellent intrasubject replicability of the AEF source estimates, Lütkenhöner and colleagues (2003) found no consistent dependency between sound frequency and AEF source location. The existence of multiple tonotopic regions in the cortex (Talavage *et al.*, 2000) is also likely to complicate the investigations of tonotopy from AEFs which reflect the compounded activity of multiple regions. Recent evidence for a cortical topographic representation of F0, however, comes from optical imaging of the primary auditory cortex of cat (Langner *et al.*, 2009) where orthogonal tonotopic and periodotopic maps were found.

The F0 has been shown to have a robust effect on the N1(m) latency. The latency of the N1(m) decreases as a function of F0 in the case of a variety of different stimulus types including pure tones (Jacobson *et al.*, 1992; Woods *et al.*, 1993; Roberts and Poeppel, 1996; Stufflebeam *et al.*, 1998; Lütkenhöner *et al.*, 2001; Seither-Preisler *et al.*, 2003), click trains (Forss *et al.*, 1993), synthesized speech sounds (Poeppel and Roberts, 1996), triangle and square waves (Roberts *et al.*, 1998), bandpass-filtered harmonic tones (Ragot and Lepaul-Ercole, 1996; Crottaz-Herbette and Ragot, 2000), and IRNs

(Krumbholz *et al.*, 2003) in the F0 range below 1 kHz. As the traveling wave on the basilar membrane is delayed for low frequency sounds exciting the apical end of the membrane relative to the high frequency sounds exciting the base of the membrane (Moore, 2004), it could be argued that the relationship between the N1(m) latency and stimulus frequency arises already from these cochlear delays. However, Seither-Preisler and colleagues (2006) compensating for the cochlear delays according to Patterson (1994), demonstrated that the trend between the N1m latency and stimulus frequency could not be accounted by the delays of the traveling wave alone. Furthermore, Roberts and Poeppel (1996) observed a U-shaped dependency between the N1m latency and stimulus frequency where the latency was shorter for middle frequencies (1000-2000 Hz) than for low (100-500 Hz) or high (3000-5000 Hz) frequencies. In contrast to this U-shaped curve, the delays in the cochlear traveling wave increase monotonically with decreasing frequency. Another possible explanation for the relationship between the N1(m) latency and stimulus F0 is that both perceptual (Moore, 2004) and auditory nerve (*e.g.*, Temchin *et al.*, 2008) thresholds show a U-shaped relationship with frequency. Therefore, the latency effect could be explained by the ear's sensitivity to different frequency ranges. This interpretation rests on the assumption that the N1(m) latency depends on the amount of peripheral activity which is indirectly supported by the results of Onishi and Davis (1968) where the N1 latency decreases with increasing sound level pressure. While this explanation seems valid for pure tones, it might not account for the latency effects observed with broadband sounds where the sound energy is distributed throughout the cochlea. Nevertheless, the peripheral auditory phenomena such as cochlear wave delays and the cochlear audiogram, that is, the ear's sensitivity to different frequencies should be taken into account when studying the cortical F0-dependent activity.

There is evidence that suggests a contribution of neural mechanisms that mediate pitch perception to the dependency between N1m latency and F0. Ragot and Crottaz (1998) presented subjects with harmonic tones with spectral comb structure where the inter-harmonic spacing was smaller than the critical bandwidth of the auditory filters which describe the frequency resolution of the auditory system (*cf.*, Shackleton and Carlyon, 1994). The perceived pitch of such sounds, with so-called unresolved harmonics, is dependent on the phase relations between the harmonics. In particular, when the odd- and even-numbered harmonics of the F0 are set into sine and into cosine phase, respectively, a pseudoperiod with a duration of 1/2F0 emerges in the sound waveform and the perceived pitch is an octave above the F0. Ragot and Crottaz (1998) observed that the latency of the N1m was shorter in the condition with unresolved harmonics in alternate phases than in the

condition with harmonics in the sine phase. Thus, the N1m latency reflected the time-domain structure and the perceived pitch of the stimulus rather than the nominal F0 or features of the power spectrum. Also in the case of stimuli composed of two sinusoids or in the case of a sinusoidally amplitude modulated sinusoid, the N1m latency was apparently determined by the perceived pitch rather than by the lowest audible partial in the sound spectrum (Roberts *et al.*, 1998).

In summary, EEG and MEG studies suggest that the activity of human auditory cortex is dependent on the F0 of periodic stimuli. Some of the key findings from these studies are, however, mutually inconsistent or may be due to factors that are not related to sound periodicity *per se*. Nevertheless, evidence from animal models (reviewed in Section 5.1) suggest that the F0 of a sound can be represented in the cortex by tuning properties of individual neurons (Bendor and Wang, 2005, 2010) or by the ensemble activity of distributed neurons (Bizley *et al.*, 2010).

5.5 Role of the auditory cortex in the processing of sound periodicity

The sensitivity to sound periodicity has been demonstrated not only in the auditory cortex (see Section 5.1) but also in sub-cortical structures of IC in the human brain (Griffiths *et al.*, 2001). Thus, the periodicity of speech sounds is likely to be extracted below the level of the auditory cortex. Nelken *et al.*, (2003) suggested that the role of the cortex in auditory processing is not primarily the extraction of sound features but rather, the introduction of multiple time-scales of integration over these features. For the case of speech periodicity, the longer time-scales of auditory processing on the cortical level would seem to be a prerequisite for the encoding of intonation, that is, F0-contours of speech. In several hemodynamic studies, the neural processing of F0-contours has been investigated by contrasting a melody-like condition with variable F0s to a condition with a fixed F0 or to a condition with a monotonic F0 pattern. These studies have indicated significant additional periodicity-specific (or melody-specific) activation in structures higher up in the hierarchy than the primary auditory cortex such as those located in the posterior superior temporal gyri and anterior temporal lobes (Griffiths *et al.*, 1998; Griffiths *et al.*, 2001; Patterson *et al.*, 2002; Hall *et al.*, 2005; von Kriegstein *et al.*, 2010). These regions may, thus, constitute the first stages in the auditory pathway that encode the intonation patterns of speech.

Besides elaborating on the subcortical representations of sound, the auditory cortex is able to tune its input by corticofugal modulation (Zhang *et al.*, 1997; Suga and Ma, 2003; Luo *et al.*, 2008). This modulation is present in the processing of several sound features including frequency, duration, echo delay, and spatial cues (Suga and Ma, 2003). Thus, it seems conceivable that the tuning to sound periodicity might also be modulated by descending projections from the cortex. However, the corticofugal modulation of frequency tuning has so far been investigated only with pure tone stimuli and, thus, the results may not be representative of the processing of the periodicity of natural broad band sounds.

In conclusion, the periodicity-sensitivity of the auditory cortical areas may be used for maintaining, elaborating, and tuning the subcortically generated representations of sound periodicity. In this view, the auditory cortex plays an important role in the encoding periodicity-related features of speech sounds.

6. Overview of the studies in the dissertation

6.1 Motivation of the studies

While significant advances in understanding the processing of sound periodicity in the auditory cortex has been gained from the studies reviewed in the previous section, the issue of generalizing the results from studies using non-speech stimuli such as iterated ripple noise and click trains to the perception of speech or even to the perception of other broad-band sounds remains problematic. Additionally, there are several speech-related issues regarding the cortical processing of periodicity that cannot be resolved based on the available data. These issues comprise the lower F0 limit of cortical sensitivity to the periodicity of speech sounds, the relationship between cortical periodicity-sensitive activity and the perceptual vocal quality of speech, the temporal integration of speech periodicity underlying cortical periodicity-sensitive activity, the contextual effects in cortical processing of periodicity, and the possibility of cortical sensitivity to the aperiodicities of speech sounds. In the series of studies presented in this thesis (Section 7), these particular issues were investigated in a systematic way by varying the periodicity-related features of vowel stimuli recorded from a human speaker.

6.2 General methods of the studies

MEG measurements and analyses: In the studies of this thesis, the cortical sensitivity to the periodicity of speech sounds was investigated primarily by comparing the cortical activity elicited by vowel sounds with variable degrees of periodicity. In particular, aperiodic versions of the vowel sounds were synthesized by replacing the periodic excitation with a random excitation

waveform. The activity elicited by the aperiodic vowel stimuli was then used as the reference against which periodicity-specific activity was contrasted. Using this approach, specific aspects of periodicity processing such as temporal integration of vowel periodicity, lower limit of pitch sensation, and contextual effects on the processing of periodicity were investigated in further detail in the studies of this thesis.

The subjects participating in the MEG studies consisted of healthy, right-handed, Finnish-speaking volunteers with average age below 28 years. The number of participants in a given study ranged from ten to fifteen and the order of presentation of the experimental conditions was counter-balanced across the subjects in each study. During MEG measurements, the subjects, instructed not to pay attention to the auditory stimuli, concentrated on watching a silent video. The subjects were also instructed to avoid eye movements and blinks during MEG data acquisition. The studies were approved by the Ethical Committee of Helsinki University Central Hospital.

In all studies, a 306-channel whole-head neuromagnetometer (Vectorview 4-D, Elekta Neuromag, Helsinki, Finland) was used for the recordings of the brain activity in a magnetically shielded booth. The data were collected with a recording bandwidth of 0.1–200 Hz and sampled at a rate of 600 Hz. At the beginning of each stimulus sequence, the head position with respect to the sensor array was determined by using head position indicator coils attached to the subject's scalp, with the locations of the coils with respect to the left and right preauricular points and the nasion having been determined prior to the measurement. On-line averaging of MEG data was synchronized to a trigger signal which was coupled to the onset of each stimulus or stimulus pair (in Study IV). The averaged epochs contained a 100-ms pre-trigger baseline and post-stimulus activation extending beyond the peak of cortical activation elicited by the auditory stimuli. In each experimental condition of the studies, 150 artifact-free averaged epochs were acquired. The epochs containing electro-oculogram values exceeding |150| μv or MEG values in the excess of |3000| fT/cm were rejected online.

The amplitudes and generator locations of the N1m (Studies 1-4) and the SF (Studies 1-2) responses were investigated with the ECD modeling technique in each hemisphere separately, with the assumption of a single dipole in a spherical volume conductor. Latency analysis was restricted to the transient N1m responses because these, unlike SF responses, have well defined peaks. The ECDs were fitted to the maximum amplitude points of the auditory N1m and SF in the data from a subset of 44 planar gradiometers in the left and in the right hemisphere separately. In order to complement the source level (ECD) analyses of AEF amplitudes (in nAm units), sensor level amplitudes (in

fT/cm units) derived from the planar gradiometers directly above the temporal cortices were inspected in Studies II and III.

Stimulus generation: The vowel stimuli used in the studies were produced semisynthetically to possess a spectral structure with realistic formant frequencies and, in the case of periodic vowel stimuli, excitation signals with highly natural-like pulseforms. The generation of the vowel stimuli consisted of the following steps: 1) recording samples of vowel sounds with a microphone from a human speaker, 2) extracting the contribution of the excitation and the articulatory filtering from the recorded speech waveform by means of inverse filtering, 3) manipulating the excitation signal independently of the effects of the articulatory filtering, and 4) applying the filter function derived from the original vowel sound to the excitation signal in order to produce a highly realistic formant pattern to the vowel stimulus. The manipulations in vowel excitation were targeted to the F0 (Study I), to the cycle-to-cycle variability in period length (Study II) and to the presence/absence of a periodic structure (all studies). In Study III, the duration of the vowel stimuli were varied.

Behavioral measures: In order to assess whether modifications in the periodic structure of vowel stimuli would affect the intelligibility of these stimuli, the subjects participating in Study I completed a vowel recognition task. In this behavioral experiment subjects were presented with vowel stimuli [a] and [e], used in the MEG studies, and were asked to indicate with a key press which of the (eight Finnish) vowels they perceived. The average percentage of correct identifications over subjects was 88.8 % which is significantly over the chance level (1/8 = 12.5%). Thus, both periodic and aperiodic stimuli could easily be identified correctly.

In Study II, the relationship between vocal jitter and the perceived roughness of the vowel stimuli was studied with a perceptual scaling experiment. The vocal quality was assessed with an anchored perceptual scaling procedure where listeners assigned a numerical scale value to each test stimulus with the help of two reference stimuli having the smallest (1%) and the largest (13%) degree of vocal jitter in the set of vowel stimuli. Each trial of the experiment consisted of three successive vowel stimuli with 400-ms silent periods between the stimuli. The first and the last stimulus were always the anchor stimuli, and the middle stimulus was selected randomly from the stimulus set as a test stimulus to be judged by the listener using a five-point rating scale. The smallest scale value (1) indicated the least degree of vocal roughness and the largest scale value (5) indicated the largest degree of perceived vocal roughness. Both vowels [a] and [e] were assessed, and the

anchor stimuli used in each trial were always matched to the test stimuli with respect to vowel identity.

7. Summaries of the publications

7.1 Cortical sensitivity to periodicity of speech sounds (Study I)

The aim of the study was to determine the relationship between cortical periodicity-specific activity and the F0 of vowel sounds in the F0 range extending from pitch to infrapitch range (limited approximately by a threshold of 20–40 Hz). To this end, subjects were presented with periodic vowel stimuli [a] and [e] with F0s in the 9–113 Hz range. The cortical activity, reflected in the N1m and SF responses, elicited by the periodic vowels was compared to the activity elicited by aperiodic vowel stimuli with matched spectral envelopes.

The results indicated sensitivity to vowel periodicity in the auditory cortex in terms of an increased amplitude of the N1m and the SF response in the condition of periodic as opposed to aperiodic stimulation (Fig. 6). The source location of the responses was, further, more anterior in the periodic than in the aperiodic condition. The periodicity-sensitivity reflected in the AEF amplitude and source location persisted for all F0 values of the vowel stimuli, even in the infrapitch range.

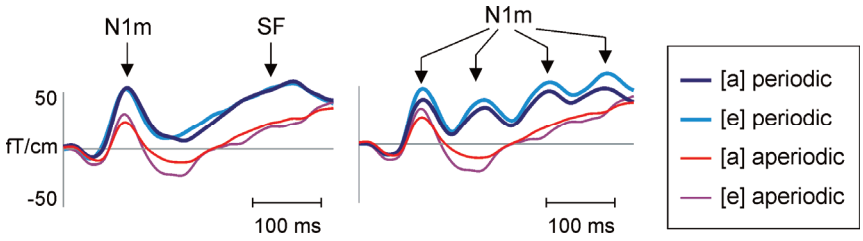

Figure 6. N1m and sustained field (SF) responses elicited by periodic and aperiodic vowel stimuli [a] and [e]. The response amplitude for both the N1m and the SF was larger in the periodic (thick line) than in the aperiodic (thin line) condition. The periodic vowel with an F0 in the pitch range (above 19 Hz) elicited a single N1m response followed by a SF (shown in the left panel). The periodic vowel with an F0 in the infrapitch range (9 Hz) elicited a series of N1m-like responses following the repetition cycle of the stimulus (shown in the right panel).

The latency of the N1m depended on the vowel F0. In the 19–113 Hz range, the latency decreased monotonically as a function of the F0. However, in the 9-Hz condition, the trend that related the N1m latency to stimulus F0 broke down and the latency was markedly shorter than that predicted by the trend. Moreover, in the 9-Hz condition, a series of N1m-like transient responses following the vowel pulsation was observed instead of the typical single N1m response at stimulus onset (Fig. 6, right). These phenomena in the N1m latency and in the AEF shape may reflect qualitative changes in the neural processing of vowel periodicity below the lower limit of pitch perception.

The amplitude of both the N1m and the SF response depended, in addition to vowel periodicity, on the formant frequencies of the vowel stimuli. That is, a larger amplitude of the AEFs was observed in the condition with vowel stimulus [e] than in the condition with vowel stimulus [a]. The dependency of the AEF amplitude on the formant frequencies of the vowel stimuli could reflect either phonemic processing of vowel identity or a more general auditory sensitivity to the spectral envelope of a sound stimulus. No interactions between vowel identity and periodicity were found in the elicitation of the N1m or the SF response, however. This suggests that the underlying generators of the AEFs represent the vowel periodicity invariantly of the spectral envelope of the stimulus.

7.2 Representation of the vocal roughness of aperiodic speech sounds in the auditory cortex (Study II)

The aim of the study was to investigate the relationships between the degree of vowel periodicity, perceived vocal roughness, and cortical periodicity-sensitive activity. The degree of vowel periodicity was manipulated by introducing increased vocal jitter that may occur, for instance, in pathological voice production. The jitter was defined as the random variability in the glottal interpulse intervals and ranged from < 1 % (the original vowel stimulus) to 13 % relative to the average interpulse interval of 9.3 ms (F0 = 107). The AEFs, as indexes of cortical activity, elicited by vowel stimuli with variable jitter were, then, compared against the AEFs elicited by aperiodic, noise-excited, vowel stimuli.

The results showed a decrease in the N1m and the SF amplitude (Fig. 7) as a function of jitter (decreased degree of vowel periodicity) of the vowel stimuli. The ratings of perceived vocal roughness derived from the behavioral evaluations of vocal quality of the vowel stimuli indicated, in turn, an increase in the perceived vocal roughness as a function of jitter. A negative correlation

between perceived vocal roughness and AEF amplitude was observed not only in the group average data but also in individual subjects as well.

The periodicity-specific enhancement in the response amplitude was larger for the SF than for the N1m response in the comparison between the periodic and the aperiodic, noise-excited, condition. When this comparison was made between the periodic and the jittered condition, the sensitivity to periodicity, conversely, appeared to be more prominent in the N1m than in the SF amplitude. These differences in the sensitivity to periodicity between the N1m and the SF could, presumably, be due to the differential contribution of the cortical onset-related activity to these responses.

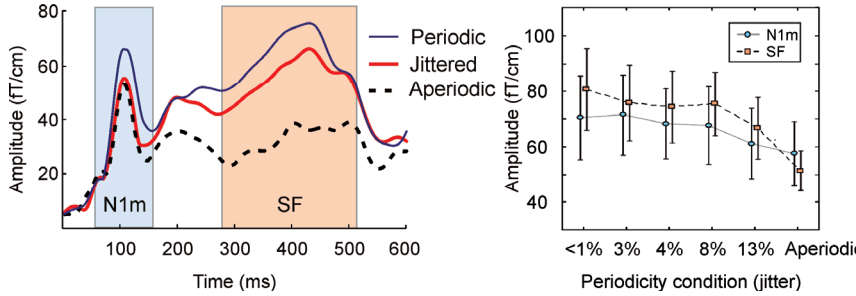

Figure 7. N1m and sustained field (SF) responses elicited by vowel stimuli characterized with differing degrees of periodicity. The amplitude of both the N1m and the SF decreased with decreasing degree of vowel periodicity. AEFs elicited by a periodic vowel, by a vowel with 13 % jitter, and by an aperiodic vowel are shown on the left. The AEF amplitude in conditions of variable jitter (< 1–13 %) and in the aperiodic condition is shown on the right.

The results are in line with the view that the degree of sound periodicity is represented in cortex by the activity of a periodicity-sensitive population. In particular, the results indicate that the relationship between the AEF amplitude and the degree of stimulus periodicity previously observed with non-speech stimuli (*e.g.*, Gutschalk *et al.*, 2007) holds for cortical activity elicited by vowel stimuli with various degrees of natural-like vocal jitter.

7.3 Temporal integration of vowel periodicity in the auditory cortex (Study III)

The aim of the study was to investigate the temporal window of integration (TWI) which underlies the cortical sensitivity to the periodicity of vowel stimuli. The duration of periodic and aperiodic vowel stimuli (F0 ≈ 100 Hz) was varied in the 10–100 ms range, and periodicity-sensitive activity was operationalized as the difference in the characteristics of the N1m response between the periodic and the aperiodic condition. The (minimal) integration window of periodicity was determined by the shortest stimulus duration at which periodicity-sensitive N1m responses could be observed.

The results indicated larger N1m amplitude in the periodic than in the aperiodic vowel condition when the duration of the vowel stimuli was at least three times the period length of the vowel, that is, 30 ms (Fig. 8). Moreover, the source location of the N1m was more anterior in the periodic than in the aperiodic condition provided that the vowel duration was at least 30 ms. Finally, the N1m latency was shorter for the periodic than for the aperiodic condition when the vowel duration was five period lengths or more.

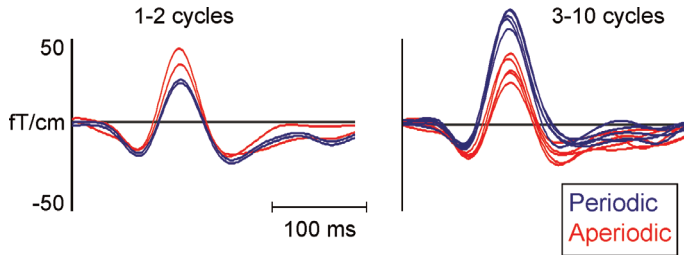

Figure 8. N1m responses elicited by periodic and aperiodic vowels of differing duration (1–10 cycles, 10–100 ms). For durations of 1–2 cycles there was no significant difference between the periodic and the aperiodic condition. For vowel durations of 3 cycles or more, the amplitude of the N1m was larger in the periodic than in the aperiodic condition.

The relationship between the N1m amplitude and vowel duration were explained with a model according to which the N1m reflects the compounded activity of periodicity-sensitive and sound energy-sensitive neural generators. The TWI of energy-sensitive N1m generators was estimated from the responses elicited by the aperiodic vowel stimuli (the periodicity-sensitive population was assumed to be inactive in this condition). The TWI of the periodicity-sensitive population was then estimated by correlating the integral of a periodicity measure, calculated from the stimulus waveform, to the

periodicity-sensitivity of the N1m amplitude. In estimating the TWI for periodicity, the changes in the N1m amplitude that could be accounted by changes in stimulus energy were controlled for by matching the stimulus intensity between the periodic and the aperiodic vowel stimuli.

The periodicity-specific enhancement of the N1m amplitude in the periodic relative to the aperiodic condition further increased somewhat when the vowel duration was increased beyond the minimal integration time of 3 stimulus cycles. Thus, as an estimate of the maximal TWI for periodicity in the generators of the N1m, a length of 5 stimulus cycles (50 ms) was obtained.

7.4 Cortical encoding of aperiodic and periodic speech sounds: evidence for distinct neural populations (Study IV)

The aim of the study was to further investigate the cortical tuning to the degree of stimulus periodicity with a special emphasis on the processing of aperiodic stimuli. The stimulus-specific adaptation (SSA) design was used and the degree of periodicity of both the probe stimuli and the adaptor stimuli were varied. The premise was that a potential stimulus-specific release from adaptation could reveal the activity of neural populations that are tuned to the presence or to the absence of a periodic structure in the probe stimulus. The paradigm also allowed the investigation of effects related to stimulus timing, which may provide insights into cortical processing of natural speech where articulation of sounds takes place at relatively rapid intervals. In the MEG experiment, both periodic and aperiodic vowels were used as probes and as adaptors which were separated by interstimulus gaps (ISG) of either 800 ms or 200 ms.

The results indicated cortical sensitivity to periodicity as indexed by the N1m response. This had a larger amplitude and a more anterior source location when elicited by periodic than when elicited by aperiodic probe stimuli. Stimulus-specific release from adaptation was, in turn, observed when the aperiodic probe was preceded by the periodic adaptor (Fig. 9). Importantly, this release from adaptation suggests the activation of a distinct population that is sensitive to vowel aperiodicity.

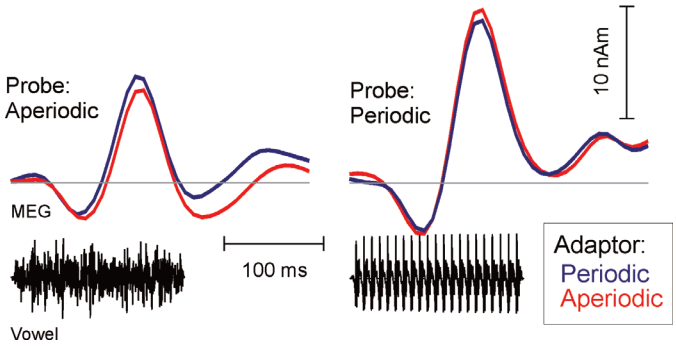

Figure 9. N1m responses elicited by periodic and aperiodic probes in a stimulus-specific adaptation paradigm. The N1m amplitude for the aperiodic probe was larger when the probe was preceded by a periodic than by an aperiodic adaptor stimulus. Thus, stimulus-specific release from adaptation was observed when the degree of periodicity of the probe differed from that of the adaptor.

The effect of ISG was manifested in the N1m latency which increased by about 20 ms when the gap was reduced from 800 ms to 200 ms. Moreover, the periodicity-specific enhancement of the N1m amplitude was abolished in the 200-ms ISG condition. However, the source location of the N1m in the periodic condition was persistently more anterior than in the aperiodic condition despite the manipulations in the ISG.

The results suggest that auditory cortex contains both aperiodicity-sensitive and periodicity-sensitive neural populations, the activity of which might tentatively be used to encode the degree of periodicity of speech sounds. The activity of the periodicity-sensitive generator of the N1m was also shown to be highly sensitive to the ISG. Finally, the periodicity-specific shift in the N1m source location, previously revealed with stimulation rates that are much slower than those found in natural speech, was observed here in the 200-ms ISG condition as well.

8. General discussion

Periodicity is an important feature in speech communication which is related to phenomena such as phonological contrasts, intonation, tonality, and voice quality. Further, sensitivity to sound periodicity in the auditory cortex has been shown in a number of EEG/MEG and hemodynamic studies of the human brain. However, the cortical processing of sound periodicity has typically been studied with non-speech stimuli which may elicit activity in the brain that differs in many ways from that elicited by speech sounds and other human vocalizations (Belin *et al.*, 2000; Gunji *et al.*, 2003; Hewson-Stoate *et al.*, 2006; Whalen *et al.*, 2006; Gutschalk *et al.*, 2011). Consequently, in order to determine the cortical mechanisms involved in the processing of the periodicity of speech, it is essential to use realistic speech sound stimuli.

In the studies of this thesis, the auditory cortical activity elicited by periodic and aperiodic vowel stimuli was measured with MEG. Using semisynthetic speech generation, key features related to vowel periodicity, namely F0 and the degree of periodicity were manipulated independently of the spectral envelope and the formant frequencies. The effects of vowel duration and preceding stimulus context on the cortical activity elicited by periodic and aperiodic vowels were investigated as well. The investigation of cortical activity focused on the N1m and SF responses, which have been previously shown to reflect stimulus-specific processing and are known to adapt to the features of the preceding stimulation (see Sections 4 and 5). By matching the periodic vowel stimuli to the aperiodic ones with respect to all other significant acoustic features, it was possible to explore the cortical activity that was specifically related to vowel periodicity. Thus, strict experimental control with carefully matched reference conditions were combined with the use of realistic periodic vowel stimuli.

The studies of the thesis revealed a sensitivity of the auditory cortex to the periodicity of speech sounds. This sensitivity to periodicity was manifested in terms of an increased amplitude and a more anterior source location of both the N1m and the SF response in conditions of periodic as opposed to aperiodic vowel stimuli. Furthermore, this periodicity-sensitivity was observed for

realistic vowel sounds with variable F0s, phonemic identities, durations, degrees of periodicity, and interstimulus gaps. Thus, the results indicate that the human auditory cortex is highly sensitive to the periodicity of speech sounds. The persistence of this sensitivity over a wide range of variations in the acoustic features of vowel stimuli suggests that the neural activity underlying the AEFs could be used to encode the periodicity of the inherently variable sounds that are used in natural speech communication.

An explanation for the observed sensitivity to periodicity of the AEFs can be given in terms of dedicated feature-selective populations in cortex. That is, a distinct cortical population is activated preferentially by periodic sounds. This hypothesis is supported by both hemodynamic (Griffiths *et al.*, 1998, 2001; Patterson *et al.*, 2002; Penagos *et al.*, 2004, Hall *et al.*, 2005, 2006; Hall and Plack, 2009; Lewis *et al.*, 2009; Garcia *et al.*, 2010; von Kriegstein *et al.*, 2010) and invasive (Schönwiesner and Zatorre, 2008; see also Griffiths *et al.*, 2010) studies of the human brain showing spatially distinct activation patterns for periodic and aperiodic stimuli. The enhancement of the amplitude of the AEFs observed for periodic relative to aperiodic stimulus conditions in the studies of the thesis and in several previous studies (Hertrich *et al.*, 2000; Alku *et al.*, 2001; Gutschalk *et al.*, 2004; Soeta *et al.*, 2005; Tiitinen *et al.*, 2005; Lütkenhöner *et al.*, 2006) is consistent with this view. In the studies of this thesis, periodicity-sensitivity was, importantly, observed for realistic vowel sounds characterized by variable stimulus features.

Compelling evidence points to the left-lateralization of speech perception (*e.g.*, Belin *et al.*, 1998; Jancke *et al.*, 2002; Phillips and Farmer, 1990; Zaehle *et al.*, 2004) which is complemented by the right-lateralized processing of the pitch of periodic sounds (Zatorre, 2001). This lateralization of function would seem to predict that the sensitivity to periodicity observed in the AEFs could be more prominent in the right than in the left hemisphere. The current results, however, indicate a symmetrical bilateral sensitivity to vowel periodicity. Such bilateral activation elicited by sound periodicity has also been observed in several other MEG (Hertrich *et al.*, 2000; Alku *et al.*, 2001; Gutschalk *et al.*, 2004; Soeta *et al.*, 2005; Tiitinen *et al.*, 2005) and hemodynamic studies (*e.g.*, Griffiths *et al.*, 1998, 2001; Patterson *et al.*, 2002; von Kriegstein *et al.*, 2010). According to the results of Patterson and colleagues (2002), it seems that the right-hemispheric dominance in the processing of sound periodicity is observed in the case of melody-like variations in the F0 but not for sounds with a fixed F0. Thus, the current results are in line with several other studies in suggesting that the periodicity-specific activity elicited by isolated steady-state auditory stimuli is distributed symmetrically across the left and the right cortical hemisphere. Further, the current thesis suggests that, despite the general tendency of left-lateralized

processing of speech sounds, a bilateral pattern of activity seems to characterize the representation of the periodicity of individual vowel sounds.

Although the cortical sensitivity to speech periodicity was observed with vowel stimuli with a wide range of F0s, durations, and degrees of periodicity, this sensitivity was modulated or abolished by strong deviations in the above features from the values that characterize normal speech sounds. These modulations effected by changes in the F0, duration, and degree of periodicity can be suggested to reflect the lower limit of pitch perception, temporal integration of periodicity, and the representation of vocal quality of jittered speech sounds, respectively. In the first study of this thesis, the changes in the N1m latency and the multiplication of the N1m response were observed in the F0 region below 19 Hz which has been previously associated with the lower limit of pitch perception (40 Hz, Ritsma, 1962; 19 Hz Guttman and Julesz, 1963; 30 Hz, Krumbholz *et al.*, 2000; 30 Hz, Pressnitzer *et al.*, 2001). In natural speech production, the degree of sound periodicity may be decreased due to voice pathologies (Lieberman, 1963; Iwata and von Leden, 1970; Murry and Doherty, 1980) or due to non-modal phonation (Henton and Bladon, 1988; Childers and Lee, 1991; Laver, 1994; Blomgren *et al.*, 1998). The results of the second study of this thesis indicate that the periodicity-sensitivity of the AEFs is correlated to both the degree of periodicity and the auditory-perceptual voice quality of vowel sounds. Finally, the detection of sound periodicity requires temporal integration (Plack and Oxenham, 2005b) where the number of cycles rather than absolute time seems to be integrated (Wiegrebe, 2001). According to the results of the third study of this thesis, the minimum duration of a vowel stimulus with an F0 of around 100 Hz for the elicitation of cortical periodicity-specific activity is three cycle lengths.

The fourth study of this thesis suggests that in addition to the periodicity-sensitive population, distinct aperiodicity-sensitive processing might be activated by aperiodic speech sounds. Although many previous non-invasive studies of the human brain have failed to reveal such aperiodicity-sensitive activity (Mäkelä *et al.*, 1988; Krumbholz *et al.*, 2003; Gutschalk *et al.*, 2004; Lütkenhöner *et al.*, 2011), the elicitation of an N1 by the onset of aperiodicity in a continuous stimulus (Martin and Boothroyd, 1999) and the hemodynamic aperiodicity-sensitive contrast in posteromedial auditory areas (von Kriegstein *et al.*, 2010) support the view of aperiodicity-sensitive activity in the auditory cortex. It may be that the observation of such aperiodicity-specific activity is conditional to using speech sound stimuli since the negative findings were obtained with non-speech stimuli including square-waves (Mäkelä *et al.*, 1988), IRN (Krumbholz *et al.*, 2003), and click-trains (Gutschalk *et al.*, 2004; Lütkenhöner *et al.*, 2011). Thus, the periodicity of speech sounds could be encoded by distinct periodicity-sensitive and aperiodicity-sensitive

populations each activated preferentially by sounds with a degree of periodicity that matches the tuning characteristics of the population.

The mapping between the AEFs measured non-invasively with MEG and the simultaneous activity of large amounts of cerebral neurons is not one-to-one. That is, similar AEFs may be produced by different patterns of neural activity. Therefore, alternative explanations to the current MEG data might be considered. That is, similar observations of periodicity-sensitive enhancement of the AEF amplitude could be possible in terms of an increased activity or synchronization in cortical populations that are activated by both periodic and aperiodic stimuli. The view of dedicated populations for encoding the periodicity and the aperiodicity of speech sounds is, however, consistent with the prevailing conception of feature-selectivity in the auditory and other sensory cortices (*cf.*, Shamma, 2001). It is also in line with recent experimental results from human hemodynamic and intracortical studies as well as with animal studies of cortical processing of periodicity. Furthermore, feature-selective populations provide a plausible explanation for the differences in the source location of the AEFs elicited by periodic and aperiodic vowels and for the stimulus-specific adaptation effects observed in the studies of this thesis.

9. Conclusions

The results of this thesis suggest that the human auditory cortex is highly sensitive to the periodicity of speech sounds. Further, they reveal the dependency of this sensitivity to communicatively significant features of speech sounds such as the fundamental frequency and the degree of periodicity. An interpretation of the current results in the light of the prevailing view of cortical feature-selectivity would suggest that the elementary features of speech sounds that are related to periodicity are encoded bilaterally in the activity of cortical periodicity- and aperiodicity-sensitive populations.

References

Ahlfors, S. P., Han, J., Lin, F.-H., Witzel, T., Belliveau, J. W., Hämäläinen, M. S., and Halgren, E. (2010). Cancellation of EEG and MEG signals generated by extended and distributed sources. *Hum. Brain Mapp*. 31:140-149.

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., and Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proc. Natl. Acad. Sci. USA*. 103:14608-14613.

Airas, M. (2008). *Methods and Studies of Laryngeal Voice Quality Analysis in Speech Production*, Doctoral thesis: TKK, Espoo, Finland.

Alku, P. (1992). Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. *Speech Commun*., 11:109-118.

Alku, P., Bäckström, T., and Vilkman, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *J. Acoust. Soc. Am*., 112:701-710.

Alku, P., Magi, C., Yrttiaho, S., Bäckström, T., and Story, B. (2009). Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering. *J. Acoust. Soc. Am*., 125:3289-3305.

Alku, P., Sivonen, P., Palomäki, K., and Tiitinen, H. (2001). The periodic structure of vowel sounds is reflected in human electromagnetic brain responses. *Neurosci. Lett*., 298:25-28.

Alku, P., Tiitinen, H., and Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clin. Neurophysiol*., 110:1329-1333.

Alku, P., and Vilkman, E. (1996). A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. *Folia Phoniatr. Logop*., 48:240-254.

Baars, B. J., and Ramsay, T. (2007). The tools: Imaging the living brain. In Baars, B. J., and Gage, N. M. (Eds.) *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, New York: Academic Press.

Baillet, S. (2010). The dowser in the fields: Searching for MEG sources. In Hansen, P., Kringelbach, M., and Salmelin, R. (Eds.) *MEG - An Introduction to Methods*, New York: Oxford University Press.

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*, San Diego: Singular (Thomson Learning).

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403:309-312.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., and Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *J. Cognit. Neurosci.*, 10:536-540.

Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436:1161-1165.

Bendor, D., and Wang, X. (2010). Neural coding of periodicity in marmoset auditory cortex. *J. Neurophysiol.*, 103:1809-1822.

Bizley, J. K., Walker, K. M., King, A. J., and Schnupp, J. W. (2010). Neural ensemble codes for stimulus periodicity in auditory cortex. *J. Neurosci.*, 30:5078-5091.

Bizley, J. K., Walker, K. M., Silverman, B. W., King, A. J., and Schnupp, J. W. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J. Neurosci.*, 29:2064-2075.

Blauert, J., and Xiang, N. (2009). *Acoustics for Engineers: Troy Lectures*, Berlin: Springer.

Bloch, B. (1941). Phonemic overlapping. *Am. Speech.*, 16:278-284.

Blomgren, M., Chen, Y., Ng, M. L., and Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *J. Acoust. Soc. Am.*, 103:2649-2658.

Bloomfield, L. (1933). *Language*, New York: Henry Holt.

Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *J. Acoust. Soc. Am.*, 44:945-950.

Campbell, N., and Mokhtari, P. (2003). Voice quality: The 4th prosodic dimension. *Proc. XVth Int. Congr. Phonet. Sci.*, Barcelona, 3:2417-2420.

Catford, J. C. (1977). *Fundamental Problems in Phonetics*, Edinburgh (UK): Edinburgh University Press.

Childers, D. G., Hicks, D. M., Moore, G. P., Eskenazi, L., and Lalwani, A. L. (1990). Electroglottography and vocal fold physiology. *J. Speech. Hear. Res.*, 33:245-254.

Childers, D. G., and Lee, C. K. (1991). Vocal quality factors: Analysis, synthesis, and perception. *J. Acoust. Soc. Am.*, 90:2394-2410.

Chuang, C. K., and Wang, W. S. (1978). Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *J. Acoust. Soc. Am.*, 64:1004-1014.

Crottaz-Herbette, S., and Ragot, R. (2000). Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clin. Neurophysiol.*, 111:1759-1766.

de Cheveigné, A. (1998). Cancellation model of pitch perception. *J. Acoust. Soc. Am.*, 103:1261-1271.

de Cheveigné, A. (2005). Pitch perception models. In Plack, C. J., Oxenham, A., Fay, R. R., and Popper, A. N. (Eds.) *Pitch: Neural Coding and Perception*, New York: Springer.

de Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111:1917-1930.

Dejonckere, P. H. (2010). Voice evaluation and respiratory function assessment. In Anniko, M., Bernal-Sprekelsen, M., Bonkowsky, V., Bradley, P., and Iurato, S. (Eds.) *Otorhinolaryngology, Head and Neck Surgery (European Manual of Medicine)*, Berlin: Springer.

Diesch, E., Eulitz, C., Hampson, S., and Ross, B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain Lang.*, 53:143-168.

Divenyi, P. L., and Robinson, A. J. (1989). Nonlinguistic auditory capabilities in aphasia. *Brain Lang.*, 37:290-326.

Duanmu, S. (2007). *The Phonology of Standard Chinese*, Oxford: Oxford University Press.

Elert, C.-C. (1972). Tonality in Swedish: Rules and a list of minimal pairs. In Firchow, E. S., Grimstad, K., Hasselmo, N., and O'Neil, W. A. (Eds.) *Studies for Einar Haugen*, The Hague: Mouton de Gruyter.

Fant, G. (1960). *Acoustic Theory of Speech Production*, The Hague: Mouton.

Fastl, H., and Zwicker, E. (2007). *Psychoacoustics: Facts and Models*, Berlin: Springer-Verlag.

Felder, V., Jönsson-Steiner, E., Eulitz, C., and Lahiri, A. (2009). Asymmetric processing of lexical tonal contrast in Swedish. *Atten. Percept. Psychophys.*, 71:1890-1899.

Flanagan, J. (1972). *Speech Analysis, Synthesis and Perception*, New York (NY): Springer-Verlag.

Forss, N., Mäkelä, J. P., McEvoy, L., Hari, R., Sidtis, J. J., and Volpe, B. T. (1993). Temporal integration and oscillatory responses of the human auditory cortex revealed by evoked magnetic fields to click trains. *Hear. Res.*, 68:89-96.

Fröhlich, M., Michaelis, D., and Strube, H. (2001). SIM - simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. *J. Acoust. Soc. Am.*, 110:479-488.

Garcia, D., Hall, D. A., and Plack, C. J. (2010). The effect of stimulus context on pitch representations in the human auditory cortex. *Neuroimage*, 51:808-816.

Gescheider, G. A. (1997). *Psychophysics: the Fundamentals*, Mahwah (NJ): Lawrence Erlbaum Associates.

Gray, H., and Lewis, W. H. (1918). *Anatomy of the Human Body*. Philadelphia: Lea & Febiger.

Griffiths, T. D. (2005). Functional imaging of pitch processing. In Plack, C. J., Oxenham, A., Fay, R. R., and Popper, A. N. (Eds.) *Pitch: Neural coding and perception*, New York: Springer.

Griffiths, T. D., Buchel, C., Frackowiak, R. S. J., and Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.*, 1:422-427.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K., Kawasaki, H., Oya, H., Patterson, R. D., Brugge, J. F., and Howard, M. A. (2010). Direct recordings of pitch responses from human auditory cortex. *Curr. Biol.*, 20:1128-1132.

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., and Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nat. Neurosci.*, 4:633-637.

Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R., and Pantev, C. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neurosci. Lett.*, 348:13-16.

Gutschalk, A., Hämäläinen, M. S., and Melcher, J. R. (2010). BOLD responses in human auditory cortex are more closely related to transient MEG responses than to sustained ones. *J. Neurophysiol.*, 103:2015-2026.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., and Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage*, 22:755-766.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., and Rupp, A. (2007). The effect of temporal context on the sustained pitch response in human auditory cortex. *Cereb. Cortex*, 17:552-561.

Gutschalk, A., and Uppenkamp, S. (2011). Sustained responses for pitch and vowels map to similar sites in human auditory cortex. *Neuroimage*, 56:1578-1587.

Guttman, N., and Julesz, B. (1963). Lower limits of auditory periodicity analysis. *J. Acoust. Soc. Am.*, 35:610.

Hahn, M. S., Teply, B. A., Stevens, M. M., Zeitels, S. M., and Langer, R. (2006). Collagen composite hydrogels for vocal fold lamina propria restoration. *Biomaterials*, 27:1104-1109.

Hall, D. A., Barrett, D. J., Akeroyd, M. A., and Summerfield, A. Q. (2005). Cortical representations of temporal structure in sound. *J. Neurophysiol.*, 94:3181-3191.

Hall, D. A., Edmondson-Jones, A. M., and Fridriksson, J. (2006). Periodicity and frequency coding in human auditory cortex. *Eur. J. Neurosci.*, 24:3601-3610.

Hall, D. A., and Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cereb. Cortex*, 19:576-585.

Halpern, A. R., and Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex*, 9:697-704.

Hari, R., Aittoniemi, K., Järvinen, M. L., Katila, T., and Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators. *Exp. Brain. Res.*, 40:237-240.

Hellström, Å., Aaltonen, O., Raimo, I., and Vilkman, E. (1994). The role of vowel quality in pitch comparison. *J. Acoust. Soc. Am.*, 96:2133-2139.

Henrich, N. (2006). Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. *Logoped. Phoniatr. Vocol.*, 31:3-14.

Henton, C., and Bladon, A. (1998). Creak as a sociophonetic marker. In Hyman, L., Li, C. N. (Eds.) *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*, London: Routledge.

Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. *J. Acoust. Soc. Am.*, 83:257-264.

Hertrich, I., Mathiak, K., Lutzenberger, W., and Ackermann, H. (2000). Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields. *Neuroreport*, 11:4017-4020.

Hess, W. J. (2008). Pitch and voicing determination of speech with an extension toward music signals. In Benesty, J., Sondhi, M. M., and Huang, Y. (Eds.) *Springer Handbook of Speech Processing*, Berlin: Springer.

Hewson-Stoate, N., Schönwiesner, M., and Krumbholz, K. (2006). Vowel processing evokes a large sustained response anterior to primary auditory cortex. *Eur. J. Neurosci.*, 24:2661-2671.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *J. Speech. Hear. Res.*, 30:448-461.

Hillenbrand, J. (1988). Perception of aperiodicities in synthetically generated voices. *J. Acoust. Soc. Am.*, 83:2361-2371.

Hirano, M., Kurita, S., and Nakashima, T. (1981). The structure of the vocal folds. In Stevens, K. N. and Hirano, M. (Eds.) *Vocal Fold Physiology*, Tokyo: University of Tokyo Press.

Hirano, M. (1981). *Clinical Examination of Voice*, Vienna: Springer-Verlag.

Hirst, D. J., and Di Cristo, A. (1998). A survey of intonation systems. In Hirst, D. J., and Di Cristo, A. (Eds.) *Intonation Systems: A Survey of Twenty Languages*, Cambridge: Cambridge University Press.

Horii, Y. (1979). Fundamental-frequency perturbation observed in sustained phonation. *J. Speech. Hear. Res.*, 22:5-19.

Horii, Y. (1980). Vocal shimmer in sustained phonation. *J. Speech. Hear. Res.*, 23:202-209.

Huettel, S. A., Song , A. W., and McCarthy, G. (2004). *Functional Magnetic Resonance Imaging*, Sunderland (MA): Sinauer Associates.

Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46:632-639.

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of signal processing in the human brain. *Rev. Mod. Phys.*, 65:413-497.

Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic-fields of the brain: Minimum norm estimates. *Med. Biol. Eng. Comput.*, 32:35-42.

ITU-T Recommendation P.800 (1996). Methods for subjective determination of transmission quality, *Int. Telecommun. Union*.

Iwata, S., and von Leden, H. (1970). Pitch perturbations in normal and pathologic voices. *Folia Phoniatr. (Basel)*, 22:413-424.

Jacobson, G. P., Lombardi, D. M., Gibbens, N. D., Ahmad, B. K., and Newman, C. W. (1992). The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear Hear.*, 13:300-306.

Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (2000). Functional specificity in right human auditory cortex for perceiving pitch direction. *Brain*, 123:155-163.

Jäncke, L., Wustenberg, T., Scheich, H., and Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, 15:733-746.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., and Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proc. Natl. Acad. Sci. USA*, 101:6809-6814.

Kaukoranta, E., Hari, R., and Lounasmaa, O. V. (1987). Responses of the human auditory cortex to vowel onset after fricative consonants. *Exp. Brain Res.*, 69:19-23.

Keidar, A. (1983). An acoustic perceptual study of vocal fry, using synthetic stimuli. *J. Acoust. Soc. Am.*, 73:S3.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., and Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *Am. J. Speech. Lang. Pathol.*, 18:124-132.

Kreiman, J., and Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *J. Acoust. Soc. Am.*, 117:2201-2211.

Kreiman, J., Gerratt, B. R., and Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *J. Acoust. Soc. Am.*, 122:2354-2364.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *J. Acoust. Soc. Am.*, 108:1170-1180.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., and Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cereb. Cortex*, 13:765-772.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*, Oxford: Blackwell.

Langner, G. (1992). Periodicity coding in the auditory system. *Hear. Res.*, 60:115-142.

Langner, G., Dinse, H. R., and Godde, B. (2009). A map of periodicity orthogonal to frequency representation in the cat auditory cortex. *Front. Integr. Neurosci.*, 3:27.

Langner, G., Sams, M., Heil, P., and Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *J. Comp. Physiol. A.*, 181:665-676.

Laver, J. (1980). *The Phonetic Description of Voice Quality*, Cambridge (UK): Cambridge University Press.

Laver, J. (1994). *Principles of Phonetics*, Cambridge (UK): Cambridge University Press.

Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., and Brefczynski-Lewis, J. A. (2009). Human cortical organization for

processing vocalizations indicates representation of harmonic structure as a signal attribute. *J. Neurosci.*, 29:2283-2296.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psych. Rev.*, 74:431-461.

Lieberman, P. (1963). Some acoustic measures of fundamental periodicity of normal and pathologic larynges. *J. Acoust. Soc. Am.*, 35:344-353.

Liikkanen, L., Tiitinen, H., Alku, P., Leino, S., Yrttiaho, S., and May, P. (2007). The right-hemispheric auditory cortex in humans is sensitive to degraded speech sounds. *Neuroreport*, 18:601-605.

Lopes da Silva, F. (2010). Electrophysiological basis of MEG signals. In Hansen, P. C., Kringelbach, M. L., and Salmelin, R. (Eds.) *MEG: An Introduction to Methods*, New York: Oxford University Press.

Lu, T., Liang, L., and Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat. Neurosci.*, 4:1131-1138.

Lu, Z.-L., Williamson, S. J., and Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science*, 258:1668-1670.

Luo, F., Wang, Q., Kashani, A., and Yan, J. (2008). Corticofugal modulation of initial sound processing in the brain. *J. Neurosci.*, 28:11615-11621.

Lütkenhöner, B. (2003). Single-dipole analyses of the N100m are not suitable for characterizing the cortical representation of pitch. *Audiol. Neurootol.*, 8:222-233.

Lütkenhöner, B., Krumbholz, K., and Seither-Preisler, A. (2003). Studies of tonotopy based on wave N100 of the auditory evoked field are problematic. *Neuroimage*, 19:935-949.

Lütkenhöner, B., Lammertmann, C., and Knecht, S. (2001). Latency of auditory evoked field deflection N100m ruled by pitch or spectrum? *Audiol. Neurootol.*, 6:263-278.

Lütkenhöner, B., Seither-Preisler, A., Krumbholz, K., and Patterson, R. D. (2011). Auditory cortex tracks the temporal regularity of sustained noisy sounds. *Hear. Res.*, 272:85-94.

Lütkenhöner, B., Seither-Preisler, A., and Seither, S. (2006). Piano tones evoke stronger magnetic fields than pure tones or noise, both in musicians and non-musicians. *Neuroimage*, 30:927-937.

Martin, B. A., and Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear.*, 20:33-44.

Martin, B. A., Sigal, A., Kurtzberg, D., and Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *J. Acoust. Soc. Am.*, 101:1585-1599.

Martin, B. A., and Stapells, D. R. (2005). Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear Hear.*, 26:195-213.

Massaro, D. W., and Chen, T. H. (2008). The motor theory of speech perception revisited. *Psych. Bull. Rev.*, 15:453-457.

May, P., and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiol.*, 47:66-122.

Mazaheri, A., and Jensen, O. (2006). Posterior alpha activity is not phase-reset by visual stimuli. *Proc. Natl. Acad. Sci. USA*, 103:2948-2952.

Michaelis, D., Fröhlich, M., and Strube, H. W. (1998). Selection and combination of acoustic features for the description of pathologic voices. *J. Acoust. Soc. Am.*, 103:1628-1639.

Miettinen, I., Alku, P., Salminen, N., May, P., and Tiitinen, H. (2011). Responsiveness of the human auditory cortex to degraded speech sounds: Reduction of amplitude resolution vs. additive noise. *Brain Res.*, 1367:298-309.

Moore, B. C. J. (2004). *An Introduction to the Psychology of Hearing*, London: Elsevier Academic Press.

Muñoz, J., Mendoza, E., Fresneda, M. D., Carballo, G., and López, P. (2003). Acoustic and perceptual indicators of normal and pathological voice. *Folia Phoniatr. Logop.*, 55:102-114.

Murakami, S., and Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *J. Physiol.*, 575:925-936.

Murray, I. R., and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J. Acoust. Soc. Am.*, 93:1097-1108.

Murry, T., and Doherty, E. T. (1980). Selected acoustic characteristics of pathologic and normal speakers. *J Speech Hear. Res.*, 23:361-369.

Mäkelä, A. M., Alku, P., May, P. J. C., Mäkinen, V., and Tiitinen, H. (2004). Cortical activity elicited by isolated vowels and diphthongs. *Neurol. Clin. Neurophysiol.*, 91:1-4.

Mäkelä, A. M., Alku, P., May, P. J. C., Mäkinen, V., and Tiitinen, H. (2005). Left-hemispheric brain activity reflects formant transitions in speech sounds. *Neuroreport*, 16:549-553.

Mäkelä, J. P. (2006). Magnetoencephalography. Auditory evoked fields. In Burkard, R. F., Eggermont, J. J., and Don, M. (Eds.) *Auditory Evoked Potentials: Basic Principles and Clinical Application*, Philadelphia: Lippincott Williams & Wilkins.

Mäkelä, J. P., Hari, R., and Leinonen, L. (1988). Magnetic responses of the human auditory-cortex to noise square-wave transitions. *Electroencephalogr. Clin. Neurophysiol.*, 69:423-430.

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: Feature detection or something else? *Biol. Cybern.*, 89:397-406.

Neuhoff, J. G. (2004). *Ecological psychoacoustics: Introduction and history*. In Neuhoff, J. G. (Ed.) *Ecological Psychoacoustics*, San Diego (CA): Elsevier.

Obleser, J., Elbert, T., Lahiri, A., and Eulitz, C. (2003a). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cogn. Brain Res.*, 15:207-213.

Obleser, J., Lahiri, A., and Eulitz, C. (2003b). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *Neuroimage*, 20:1839-1847.

Obleser, J., Lahiri, A., and Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cogn. Neurosci.*, 16:31-39.

Onishi, S., and Davis, H. (1968). Effects of duration and rise time of tone bursts on evoked potentials. *J. Acoust. Soc. Am.*, 44:582-591.

Pantev, C., Eulitz, C., Elbert, T., and Hoke, M. (1994). The auditory evoked sustained field: Origin and frequency dependence. *Electroencephalogr. Clin. Neurophysiol.*, 90:82-90.

Pantev, C., Hoke, M., Lütkenhöner, B., and Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science*, 246:486-488.

Parkkonen, L., Fujiki, N., and Mäkelä, J. P. (2009). Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. *Hum. Brain Mapp.*, 30:1772-1782.

Patel, A. D., Peretz, I., Tramo, M., and Labrecque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain Lang.*, 61:123-144.

Patterson, R. D. (1994). The sound of a sinusoid: Spectral models. *J. Acoust. Soc. Am.*, 96:1409-1418.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *J. Acoust. Soc. Am.*, 100:3286-3294.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36:767-776.

Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.*, 24:6810-6815.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport*, 10:3979-3984.

Phillips, D. P., and Farmer, M. E. (1990). Acquired word deafness, and the temporal grain of sound representation in the primary auditory cortex. *Behav. Brain Res.*, 40:85-94.

Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, Needham Heights (MA): Allyn & Bacon.

Picton, T. W., Woods, D. L., and Proulx, G. B. (1978a). Human auditory sustained potentials: I. The nature of the response. *Electroencephalogr. Clin. Neurophysiol.*, 45:186-197.

Picton, T. W., Woods, D. L., and Proulx, G. B. (1978b). Human auditory sustained potentials: II. Stimulus relationships. *Electroencephalogr. Clin. Neurophysiol.*, 45:198-210.

Plack, C. J., and Oxenham, A. J. (2005a). The present and future of pitch. In Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (Eds.) *Pitch: Neural Coding and Perception*, New York: Springer.

Plack, C. J., and Oxenham, A. J. (2005b). The psychophysics of pitch. In Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (Eds.) *Pitch: Neural Coding and Perception*, New York: Springer.

Poeppel, D., and Roberts, T. P. L. (1996). Effects of vowel pitch and task demands on latency and amplitude of the auditory evoked M100. *Brain Cogn.*, 32:156-158.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (2001). The lower limit of melodic pitch. *J. Acoust. Soc. Am.*, 109:2074-2084.

Rabiner, L. R. (1977). On the use of autocorrelation analysis for pitch detection. *IEEE Trans. Acoust. Speech Signal Process.*, 25:24-33.

Rabiner, L. R., and Schafer, R. W. (2007). Introduction to digital speech processing. *Foundations and Trends in Signal Processing*, 1:1-194.

Ragot, R., and Crottaz, S. (1998). A dual mechanism for sound pitch perception: New evidence from brain electrophysiology. *Neuroreport*, 9:3123-3127.

Ragot, R., and Lepaul-Ercole, R. (1996). Brain potentials as objective indexes of auditory pitch extraction from harmonics. *Neuroreport*, 7:905-909.

Ritsma, R. J. (1962). Existence region of the tonal residue. I. *J. Acoust. Soc. Am.*, 34:1224-1229.

Roberts, T., Ferrari, P., and Poeppel, D. (1998). Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. *Neuroreport*, 9:3256-3269.

Roberts, T. P., Flagg, E. J., and Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport*, 15:1679-1682.

Roberts, T. P. L., and Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport*, 7:1138-1140.

Robin, D. A., Tranel, D., and Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain Lang.*, 39:539-555.

Robinson, K., and Patterson, R. D. (1995). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *J. Acoust. Soc. Am.*, 98:1858-1865.

Salmelin, R. (2010). Multi-dipole modeling in MEG. In Hansen, P. C., Kringelbach, M. L., and Salmelin, R. (Eds.) *MEG. An Introduction to Methods*, New York: Oxford University Press.

Salminen, N. H., May, P. J., Alku, P., and Tiitinen, H. (2009). A population rate code of auditory space in the human cortex. *PLoS ONE*, 26:e7600.

Salminen, N. H., Tiitinen, H., Yrttiaho, S., and May, P. J. (2010). The neural code for interaural time difference in human auditory cortex. *J. Acoust. Soc. Am.*, 127:EL60-EL65.

Schaeffler, F. (2005). *Phonological Quantity in Swedish Dialects*, Doctoral thesis: Umeå University, Umeå, Sweden.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleeck, S., Stippich, C., and Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat. Neurosci.*, 8:1241-1247.

Schönwiesner, M., and Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Exp. Brain Res.*, 187:97-105.

Seither-Preisler, A., Krumbholz, K., and Lütkenhöner, B. (2003). Sensitivity of the neuromagnetic 100m deflection to spectral bandwidth: A function of the auditory periphery? *Audiol. Neurootol.*, 8:322-337.

Seither-Preisler, A., Patterson, R., Krumbholz, K., Seither, S., and Lütkenhöner, B. (2006). Evidence of pitch processing in the N100m component of the auditory evoked field. *Hear. Res.*, 213:88-98.

Semal, C., and Demany, L. (1990). The upper limit of "musical" pitch. *Music Percept.*, 8:165-176.

Shackleton, T. M., and Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.*, 95:3529-3540.

Shah, A. S., Bressler, S. L., Knuth, K. H., Ding, M., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cereb. Cortex*, 14:476-483.

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends Cogn. Sci.*, 5:340-348.

Sidtis, J. J., and Volpe, B. T. (1988). Selective loss of complex-pitch or speech discrimination after unilateral lesion. *Brain Lang.*, 34:235-245.

Soeta, Y., Nakagawa, S., and Tonoike, M. (2005). Auditory evoked magnetic fields in relation to iterated rippled noise. *Hear. Res.*, 205:256-261.

Story, B. H., Titze, I. R., and Hoffman, E. A. (2001). The relationship of vocal tract shape to three voice qualities. *J. Acoust. Soc. Am.*, 109:1651-1667.

Strube, H. (1974). Determination of the instant of glottal closure from the speech wave. *J. Acoust. Soc. Am.*, 56:1625-1629.

Stufflebeam, S. M., Poeppel, D., Rowley, H. A., and Roberts, T. P. L. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *Neuroreport*, 9:91-94.

Suga, N., and Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat. Rev. Neurosci.*, 4:783-794.

Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., and Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear. Res.*, 150:225-244.

Temchin, A. N., Rich, N. C., and Ruggero, M. A. (2008). Threshold tuning curves of chinchilla auditory-nerve fibers. I. Dependence on characteristic frequency and relation to the magnitudes of cochlear vibrations. *J. Neurophysiol.*, 100:2889-2898.

Tiitinen, H., Mäkelä, A. M., Mäkinen, V., May, P. J. C., and Alku, P. (2005). Disentangling the effects of phonation and articulation: Hemispheric asymmetries in the auditory N1m response of the human brain. *BMC Neurosci.*, 6:62-70.

Titze, I. R. (1994). *Principles of Voice Production*, Upper Saddle River (NJ): Prentice Hall.

Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., and Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: Auditory cortex. *Ann. N.Y. Acad. Sci.*, 1060:148-174.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J. Neurosci.*, 24:10440-10453.

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.*, 6:391-398.

Uutela, K., Hämäläinen, M., and Somersalo, E. (1999). Visualization of magnetoencephalographic data using minimum current estimates. *Neuroimage*, 10:173-180.

Vainio, M., Järvikivi, J., Aalto, D., and Suni, A. (2010). Phonetic tone signals phonological quantity and word structure. *J. Acoust. Soc. Am.*, 128:1313-1321.

Ward, W. D. (1954). Subjective musical pitch. *J. Acoust. Soc. Am.*, 26:269-380.

Welham, N. V. (2009). Clinical voice evaluation. In Aronson, A. E., and Bless, D. M. (Eds.) *Clinical Voice Disorders*, New York: Thieme Medical Publishers.

Wendahl, R. W. (1966). Some parameters of auditory roughness. *Folia Phoniatr. (Basel)*, 18:26-32.

Whalen, D. H., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., Lai, S., Mencl, W. E., Fulbright, R. K., Constable, R. T., and Liberman, A. M. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *J. Acoust. Soc. Am.*, 119:575-581.

Whitehead, R. L., Metz, D. E., and Whitehead, B. H. (1984). Vibratory patterns of the vocal folds during pulse register phonation. *J. Acoust. Soc. Am.*, 75:1293-1297.

Wiegrebe, L. (2001). Searching for the time constant of neural pitch extraction. *J. Acoust. Soc. Am.*, 109:1082-1091.

Winter, I. M. (2005). The neurophysiology of pitch. In Plack, C. J., Oxenham, A., Fay, R. R., and Popper, A. N. (Eds.) *Pitch: Neural Coding and Perception*, New York: Springer.

Vipperla, R., Renals, S., and Frankel, J. (2010). Ageing voices: The effect of changes in voice parameters on ASR performance. *EURASIP J. Audio Speech Music Proc.*, Article ID 525783.

Virtanen, J., Ahveninen, J., Ilmoniemi, R. J., Näätänen, R., and Pekkonen, E. (1998). Replicability of MEG and EEG measures of the auditory N1/N1m-response. *Electroencephalogr. Clin. Neurophysiol.* 108:291-298.

von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Kiebel, S. J., and Griffiths, T. D. (2010). How the human brain recognizes speech in the context of changing speakers. *J. Neurosci.*, 30:629-638.

von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., and Griffiths, T. D. (2006). Processing the acoustic effect of size in speech sounds. *Neuroimage*, 32:368-375.

Woods, D., Alain, C., Covarrubias, D., and Zaidel, O. (1993). Frequency related differences in the speed of human auditory processing. *Hear. Res.*, 66:46-52.

Yip, M. (2002). *Tone*, Cambridge (UK): Cambridge University Press.

Yost, W. A. (1978). Strength of the pitches associated with rippled noise. *J. Acoust. Soc. Am.*, 64:485-492.

Yost, W. A. (1979). Models of the pitch and pitch strength of rippled noise. *J. Acoust. Soc. Am.*, 66:400-411.

Yost, W. A. (1996). Pitch of iterated rippled noise. *J. Acoust. Soc. Am.*, 100:511-518.

Yost, W. A., Hill, R., Perez-Falcon, T. (1978). Pitch and pitch discrimination of broadband signals with rippled power spectra. *J. Acoust. Soc. Am.*, 63:1166-1173.

Zaehle, T., Wustenberg, T., Meyer, M., and Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fMRI study. *Eur. J. Neurosci.*, 20:2447-2456.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.*, 84:566-572.

Zatorre, R. J. (2001). Neural specializations for tonal processing. *Ann. N.Y. Acad. Sci.*, 930:193-210.

Zatorre, R. J., Evans, A. C., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.*, 14:1908-1919.

Zhang, Y., and Suga, N. (1997). Corticofugal amplification of subcortical responses to single tone stimuli in the mustached bat. *J. Neurophysiol.*, 78:3489-3492.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc. Am.*, 39:151-168.

Aalto-DD 144/2011

ISBN 978-952-60-4443-9
ISBN 978-952-60-4444-6 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

**DOCTORAL
DISSERTATIONS**